38

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 643 778 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
31.05.2000 Bulletin 2000/22

(21) Application number: **93908732.6**

(22) Date of filing: **30.03.1993**

(51) Int. Cl.[7]: **C12Q 1/70**, C07K 2/00, G01N 33/53

(86) International application number:
PCT/US93/03127

(87) International publication number:
WO 93/20242 (14.10.1993 Gazette 1993/25)

(54) **ENCODED COMBINATORIAL CHEMICAL LIBRARIES**
KODIERTE KOMBINATORISCHE CHEMISCHE BIBLIOTHEKEN
BIBLIOTHEQUES CHIMIQUES COMBINATOIRES CODEES

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **30.03.1992 US 860445**

(43) Date of publication of application:
22.03.1995 Bulletin 1995/12

(73) Proprietor:
THE SCRIPPS RESEARCH INSTITUTE
La Jolla, CA 92037 (US)

(72) Inventors:
- **LERNER, Richard**
  La Jolla, CA 92037 (US)
- **JANDA, Kim**
  San Diego, CA 92107 (US)
- **BRENNER, Sydney**
  Edwards Passage, Cambridge CB2 3PJ (GB)
- **NIELSEN, John**
  San Diego, CA 92122 (US)

(74) Representative:
Fisher, Adrian John
CARPMAELS & RANSFORD
43 Bloomsbury Square
London WC1A 2RA (GB)

(56) References cited:

EP-A- 0 323 152	WO-A-90/00626
WO-A-92/00091	WO-A-92/09300
WO-A-93/06121	WO-A-93/09668
US-A- 4 923 901	US-A- 4 965 188

- **SCIENCE, vol.251, 15 February 1991 pages 767 - 778 S.P.A.FODOER ET AL. 'Light-Directed, Spatially Adressable Parallel Chemical Synthesis.'**
- **Nature, Volume 354, issued 07 November 1991, LAM et al., "A New Type of Synthetic Peptide Library for Identifying Ligand-Binding", pages 82-84, see entire document.**
- **P.N.A.S., Volume 81, issued 1984, GEYSEN et al., "Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid", pages 3998-4002, see Abstract and Materials and Methods, (Synthesis of Peptides).**
- **Tetrahedron Letters, Volume 26, issued 1985, HAYES et al., "High-Yield Synthesis of Oligoribonucleotides Using O-Nitrobenzyl Protection of 2'-Hydroxyls", pages 2407-2410, see page 2407, Summary.**
- **Nature, Volume 354, issued 07 November 1991, HOUGHTEN et al., "Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery", pages 84-86, see page 84.**

EP 0 643 778 B1

Printed by Xerox (UK) Business Services
2.16.7 (HRS)/3.6

Description

Technical Field

[0001] The present invention relates to encoded chemical libraries that contain repertoires of oligopeptide/oligonucleotide conjugates defining a diversity of biological structures, and methods for using the libraries.

Background

[0002] There is an increasing need to find new molecules which can effectively modulate a wide range of biological processes, for applications in medicine and agriculture. A standard way for searching for novel bioactive chemicals is to screen collections of natural materials, such as fermentation broths or plant extracts, or libraries of synthesized molecules using assays which can range in complexity from simple binding reactions to elaborate physiological preparations. The screens often only provide leads which then require further improvement either by empirical methods or by chemical design. The process it time-consuming and costly but it is unlikely to be totally replaced by rational methods even when they are based on detailed knowledge of the chemical structure of the target molecules. Thus, what we might call "irrational drug design" - the process of selecting the right molecules from large ensembles or repertoires - requires continual improvement both in the generation of repertoires and in the methods of selection.

[0003] Recently there have been several developments in using peptides or nucleotides to provide libraries of compounds for lead discovery. The methods were originally developed to speed up the determination of epitopes recognized by monoclonal antibodies. For example, the standard serial process of stepwise search of synthetic peptides now encompasses a variety of highly sophisticated methods in which large arrays of peptides are synthesized in parallel and screened with acceptor molecules labelled with fluorescent or other reporter groups. The sequence of any effective peptide can be decoded from its address in the array. See for example Geysen et al., Proc.Natl.Acad.Sci.USA, 81:3998-4002 (1984); Maeji et al., J.Immunol.Met., 146:83-90 (1992); and Fodor et al., Science, 251: 767-775 (1991).

[0004] In another approach, Lam et. al., Nature, 354:82-84 (1991) describes combinatorial libraries of peptides that are synthesized on resin beads such that each resin bead contains about 20 pmoles of the same peptide. The beads are screened with labelled acceptor molecules and those with bound acceptor are searched for by visual inspection, physically removed, and the peptide identified by direct sequence analysis. In principle, this method could be used with other chemical entities but it requires sensitive methods for sequence determination.

[0005] A different method of solving the problem of identification in a combinatorial peptide library is used by Houghten et al., Nature, 354:84-86 (1991). For hexapeptides of the 20 natural amino acids, 400 separate libraries are synthesized, each with the first two amino acids fixed and the remaining four positions occupied by all possible combinations. An assay, based on competition for binding or other activity, is then used to find the library with an active peptide. Then twenty new libraries are synthesized and assayed to determine the effective amino acid in the third position, and the process is reiterated in this fashion until the active hexapeptide is defined. This is analogous to the method used in searching a dictionary; the peptide is decoded by construction using a series of sieves or buckets and this makes the search logarithmic.

[0006] A very powerful biological method has recently been described in which the library of peptides is presented on the surface of a bacteriophage such that each phage has an individual peptide and contains the DNA sequence specifying it. The library is made by synthesizing a repertoire of random oligonucleotides to generate all combinations, followed by their insertion into a phage vector. Each of the sequences is cloned in one phage and the relevant peptide can be selected by finding those that bind to the particular target. The phages recovered in this way can be amplified and the selection repeated. The sequence of the peptide is decoded by sequencing the DNA. See for example Cwirla et al., Proc.Natl.Acad.Sci.USA, 87:6378-6382 (1990); Scott et al., Science, 249:386-390 (1990); and Devlin et al., Science, 249:404-406 (1990).

[0007] Another "genetic" method has been described where the libraries are the synthetic oligonucleotides themselves wherein active oligonucleotide molecules are selected by binding to an acceptor and are then amplified by the polymerase chain reaction (PCR). PCR allows serial enrichment and the structure of the active molecules is then decoded by DNA sequencing on clones generated from the PCR products. The repertoire is limited to nucleotides and the natural pyrimidine and purine bases or those modifications that preserve specific Watson-Crick pairing and can be copied by polymerases.

[0008] The main advantages of the genetic methods reside in the capacity for cloning and amplification of DNA sequences, which allows enrichment by serial selection and provides a facile method for decoding the structure of active molecules. However, the genetic repertoires are restricted to nucleotides and peptides composed of natural amino acids and a more extensive chemical repertoire is required to populate the entire universe of binding sites. In contrast, chemical methods can provide limitless repertoires but they lack the capacity for serial enrichment and there are difficulties in discovering the structures of selected active molecules.

Brief Summary of the Invention

[0009] The present invention provides a way of combining the virtues of both of the chemical and genetic methods summarized above through the construction of encoded combinatorial chemical libraries, in which each of a collection of polypeptides is labelled by an appended "genetic" tag, itself constructed by chemical synthesis, to provide a "retrogenetic" way of specifying each polypeptide. As used herein, the term "polypeptide" includes glycopolypeptides.

[0010] In outline, two alternating parallel combinatorial syntheses are performed so that the genetic tag is chemically linked to the polypeptide being synthesized; in each case, the addition of one amino acid residue to the structure is followed by the addition of an oligonucleotide sequence, which is defined to "code" for that amino acid residue, i.e., to function as an identifier for the structure of the amino acid residue. The library is built up by the repetition of this process after pooling and division.

[0011] Active molecules are selected from the library so produced by binding to a preselected biological molecule of interest or by identifying a species in the library having a desired activity, including binding, activation, chemical catalysis and the like. Thereafter, the identity of the active molecule is determined by reading the genetic tag, i.e., the identifier oligonucleotide sequence. In one embodiment, amplified copies of their retrogenetic tags can be obtained by the polymerase chain reaction.

[0012] The strands of the amplified copies with the appropriate polarity can then be used to enrich for a subset of the library by hybridization with the matching tags and the process can then be repeated on this subset. Thus serial enrichment is achieved by a process of purification exploiting linkage to a nucleotide sequence which can be amplified. Finally, the structure of the polypeptide is decoded by cloning and sequencing the products of the PCR reaction.

[0013] The present invention therefore provides a novel method for identifying a polypeptide having a preselected binding or catalysis activity through the use of a library of bifunctional molecules that provides a rich source of chemical diversity. The library is used to identify chemical structures (structural motifs) that interact with preselected biological molecules.

[0014] Thus, in one embodiment, the invention contemplates a bifunctional molecule according to claim 1.

[0015] In another embodiment, the invention contemplates a library comprising a plurality of species of bifunctional molecules, thereby forming a repertoire of chemical diversity.

[0016] Another embodiment contemplates a method for identifying a polypeptide that participates in a preselected chemical or biochemical interaction with a biologically active molecule, such as binding or catalysis, where the chemical structure is present in the library of bifunctional molecules according to this invention. The method comprises the steps of:

a) admixing in solution the library of bifunctional molecules with the biologically active molecule under binding conditions for a time period sufficient to form a binding reaction complex;
b) isolating the complex formed in step (a); and
c) determining the nucleotide sequence of the polymer identifier oligonucleotide in the isolated complex and thereby identifying the polypeptide that participated in the preselected binding interaction.

[0017] The invention also contemplates a method as claimed in claim 14 for preparing a library according to this invention.

[0018] In a related embodiment, the invention describes a bifunctional solid support for synthesizing an oligopeptide/ oligonucleotide conjugate comprising: (1) a solid support, the solid support being of a type which is dispersible in aqueous solution, (2) a first linkage unit coupled to the solid support, (3) a second linkage unit coupled to the first linkage unit, and (4) a bifunctional unit coupled to the second linkage unit, wherein the bifunctional unit having a first leaving group employable for oligopeptide synthesis and a second leaving group employable for oligonucleotide synthesis, wherein the first leaving group is N-FMOC or its functional equivalent, the second leaving group is O-DMT or its functional equivalent, wherein the second linkage unit is coupled to the first linkage unit by means of a bond cleavable by exposure to concentrated aqueous ammonia, and wherein the solid support, the first linkage unit, the second linkage unit, the cleavable bond, and the bifunctional unit, exclusive of the first and second leaving groups, each being substantially chemically unreactive to conditions employed by oligopeptide synthetic protocols using FMOC leaving groups and conditions employed by oligonucleotide synthetic protocols using O-DMT leaving groups.

[0019] Further contemplated is an element of a library of oligopeptide/oligonucleotide conjugates comprising: (1) a solid support being of a type which is dispersible in aqueous solution, (2) a first linkage unit coupled to the solid support, (3) a second linkage unit coupled to the first linkage unit, (4) a bifunctional unit coupled to the second linkage unit, (5) an oligopeptide attached to the bifunctional unit, and (6) an oligonucleotide attached to the bifunctional unit.

Brief Description of the Drawings

[0020] In the drawings, forming a portion of this disclosure:

Figure 1 illustrates a scheme for the restriction endonuclease cleavage of a PCR amplification product derived from a bifunctional molecule of this invention (Step 1), and the subsequent addition of biotin to the cleaved PCR product (Step 2).
Figure 2 illustrates the first three steps in the process of producing a library of bifunctional molecules according to the method described in Example 9.
Figure 3 illustrates the structure of a preferred bifunctional linker-support molecule based on a controlled pore glass (CPG) support designated "bf-CPG" that is useful in practicing the present invention and is described in Example 3B.

Detailed Description of the Invention

A. Encoded Combinatorial Chemical Libraries

[0021] An encoded combinatorial chemical library is a composition comprising a plurality of species of bifunctional molecules that each define a different chemical structure and that each contain a unique identifier oligonucleotide whose nucleotide sequence defines the corresponding chemical structure.

1. Bifunctional Molecules

[0022] A bifunctional molecule is the basic unit in a library of this invention, and combines the elements of a polypeptide comprised of a series of amino acid residues, and a code for identifying the structure of the polypeptide.
[0023] Thus, a bifunctional molecule can be represented by the formula A-B-C, where A is a polypeptide, B is a linker molecule operatively linked to A and C, and C is an identifier oligonucleotide comprising a sequence of nucleotides that identifies the structure of polypeptide moiety A.

a. Chemical Polymers

[0024] A chemical moiety in a bifunctional molecule of this invention is represented by A in the above formula A-B-C and is a polypeptide comprising a linear series of amino acid residues represented by the formula X_n, wherein X is a single amino acid residue in polypeptide A and n is a position identifier for X in polypeptide A. n has the value of 1+i where i is an integer from 0 to 10, such that when n is 1, X is located most proximal to the linker (B).
[0025] Although the length of the polypeptide can vary, defined by n, practical library size limitations arise if there is a large alphabet size as discussed further herein. n is therefore an integer from 4 to 50.
[0026] The chemical unit X can be selected to form a region of a natural protein or can be a non-natural polypeptide, can be comprised of natural D-amino acids, or can be comprised of non-natural amino acids or mixtures of natural and non-natural amino acids. The non-natural combinations provide for the identification of useful and unique structural motifs involved in biological interactions.
[0027] Non-natural amino acids include modified amino acids and L-amino acids, stereoisomer of D-amino acids, and any other compounds that can form an amide or pseudo-amide linkage.
[0028] The amino acid residues described herein are preferred to be in the "L" isomeric form. "H" refers to the free amino group present at the amino terminus of a polypeptide. "OH" refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, J. Biol. Chem., 243:3552-59 (1969) and adopted at 37 C.F.R. §1.822(b)(2)), abbreviations for amino acid residues are shown in the following Table of Correspondence:

TABLE OF CORRESPONDENCE		
SYMBOL		AMINO ACID
1-Letter	3-Letter	
Y	Tyr	tyrosine

(continued)

TABLE OF CORRESPONDENCE		
SYMBOL		AMINO ACID
1-Letter	3-Letter	
G	Gly	glycine
F	Phe	phenylalanine
M	Met	methionine
A	Ala	alanine
S	Ser	serine
I	Ile	isoleucine
L	Leu	leucine
T	Thr	threonine
V	Val	valine
P	Pro	proline
K	Lys	lysine
H	His	histidine
Q	Gln	glutamine
E	Glu	glutamic acid
W	Trp	tryptophan
R	Arg	arginine
D	Asp	aspartic acid
N	Asn	asparagine
C	Cys	cysteine

[0029] The phrase "amino acid residue" is broadly defined to include the amino acids listed in the Table of Correspondence and modified and unusual amino acids, such as those listed in 37 C.F.R. §1.822(b)(4), and incorporated herein by reference.

[0030] The polymer defined by chemical moiety A can therefore contain any polymer backbone modifications that provide increased chemical diversity. In building of a polypeptide system as exemplary, a variety of modifications are contemplated, including the following backbone structures: -NHN(R)CO-, -NHB(R)CO-, -NHC(RR')CO-, -NHC(=CHR)CO-, $-NHC_6H_4CO-$, $-NHCH_2CHRCO-$, $-NHCHRCH_2CO-$, and lactam structures.

[0031] In addition, amide bond modifications are contemplated including $-COCH_2-$, -COS-, -CONR, -COO-, -CSNH-, $-CH_2NH-$, $-CH_2CH_2-$, $-CH_2S-$, $-CH_2SO-$, $-CH_2SO_2-$, $-CH(CH_3)S-$, -CH=CH-, -NHCO-, -NHCONH-, -CONHO-, $-C(=CH_2)CH_2-$, $-PO_2^-NH-$, $-PO_2^-CH_2-$, $-PO_2^-CH_2N^+-$, and $-SO_2NH^--$.

b. Polymer Identifier Oligonucleotide

[0032] An identifier oligonucleotide in a bifunctional molecule of this invention is represented by C in the above formula A-B-C and is an oligonucleotide having a sequence represented by the formula Z_n, wherein Z is a sequence of 2 to 10 nucleotides within oligonucleotide C that identifies the chemical unit X at position n.

[0033] For example, a bifunctional molecule can be represented by the formula:

$$X_4X_3X_2X_1\text{-B-}Z_1Z_2Z_3Z_4.$$

In this example, the sequence of oligonucleotides Z_1, Z_2, Z_3 and Z_4 identifies the structure of amino acid residues X_1, X_2, X_3 and X_4, respectively. Thus, there is a correspondence in the identifier sequence between an amino acid residue X at position n and the unit identifier oligonucleotide Z at position n.

[0034] The length of a unit identifier oligonucleotide can vary depending on the complexity of the library, the number

of different amino acid residues to be uniquely identified, and other considerations relating to requirements for uniqueness of oligonucleotides such as hybridization and polymerase chain reaction fidelity. The length is from 2 to 10 nucleotides.

[0035] Insofar as adenosine (A), guanosine (G), thymidine (T) and cytidine (C) represent the typical choices of deoxynucleotides for inclusion in a unit identifier oligonucleotide, A, G, T and C form a representative "alphabet" used to "spell" out a unit identifier oligonucleotide's sequence. Other nucleotides or nucleotide analogs can be utilized in addition to or in place of the above four nucleotides, so long as they have the ability to form Watson-Crick pairs and be replicated by DNA polymerases in a PCR reaction. However, the nucleotides A, G, T and C are preferred.

[0036] For the design of the code in the identifier oligonucleotide, it is essential to chose a coding representation such that no significant part of the oligonucleotide sequence can occur in another unrelated combination by chance or otherwise during the manipulations of a bifunctional molecule in the library.

[0037] For example, consider a library where Z is a trinucleotide whose sequence defines a unique amino acid residue X. Because the methods of this invention provide for all combinations and permutations of an alphabet of chemical units, it is possible for two different unit identifier oligonucleotide sequences to have closely related sequences that differ by only a frame shift and therefore are not easily distinguishable by hybridization or sequencing unless the frame is clear.

[0038] Other sources of misreading of a unit identifier oligonucleotide can arise. For example, mismatch in DNA hybridization, transcription errors during a primer extension reaction to amplify or sequence the identifier oligonucleotide, and the like errors can occur during a manipulation of a bifunctional molecule.

[0039] The invention contemplates a variety of means to reduce the possibility of error in reading the identifier oligonucleotide, such as to use longer nucleotide lengths for a unit identifier nucleotide sequence as to reduce the similarity between unit identifier nucleotide sequences. Typical lengths depend on the size of the alphabet of chemical units.

[0040] A representative system useful for eliminating read errors due to frame shift or mutation is a code developed as a theoretical alternative to the genetic code and is known as the commaless genetic code. Crick et al., Proc. Natl. Acad. Sci. USA 43:416-421 (1957).

[0041] A convenient unit identifier nucleotide sequence is the well known genetic code using triplet codons or the use of a commaless code as described by Crick et al., supra. The invention need not be limited by the translation afforded between the triplet codon of the genetic code and the natural amino acids; other systems of correspondence can be assigned.

[0042] A typical and exemplary unit identifier nucleotide sequence is based on the commaless code described in the Examples, and having a length of six nucleotides (hexanucleotide) per chemical unit to provide a length that assures complementary hybridization.

[0043] Preferably, an identifier oligonucleotide has at least 15 nucleotides in the tag (coding) region for effective hybridization. In addition, considerations of the complexity of the library, the size of the alphabet of chemical units, and the length of the polymer length of the chemical moiety all contribute to length of the identifier oligonucleotide as discussed in more detail herein.

[0044] In a preferred embodiment, an identifier oligonucleotide C has a nucleotide sequence according to the formula P1-$(Z_n)_a$-P2, where P1 and P2 are nucleotide sequences that provide polymerase chain reaction (PCR) primer binding sites adapted to amplify the polymer identifier oligonucleotide. The requirements for PCR primer binding sites are generally well known in the art, but are designed to allow a PCR amplification product (a PCR-amplified duplex DNA fragment) to be formed that contains the polymer identifier oligonucleotide sequences.

[0045] The presence of the two PCR primer binding sites, P1 and P2, flanking the identifier oligonucleotide sequence $(Z_n)_a$ provides a means to produce a PCR-amplified duplex DNA fragment derived from the bifunctional molecule using PCR. This design is useful to allow the amplification of the tag sequence present on a particular bifunctional molecule for cloning and sequencing purposes in the process of reading the identifier code to determine the structure of the polypeptide in the bifunctional molecule.

[0046] More preferred is a bifunctional molecule where one or both of the nucleotide sequences P1 and P2 are designed to contain a means for removing the PCR primer binding sites from the identifier oligonucleotide sequences. Removal of the flanking P1 and P2 sequences is desirable so that their sequences do not contribute to a subsequent hybridization reaction. Preferred means for removing the PCR primer binding sites from a PCR amplification product is in the form of a restriction endonuclease site within the PCR-amplified duplex DNA fragment.

[0047] Restriction endonucleases are well known in the art and are enzymes that recognize specific lengths of duplex DNA and cleave the DNA in a sequence-specific manner.

[0048] Preferably, the restriction endonuclease sites should be positioned proximal to $(Z_n)_a$ relative to the PCR primer binding sites to maximize the amount of P1 and P2 that is removed upon treating a bifunctional molecule to the specific restriction endonuclease. More preferably, P1 and P2 each are adapted to form a restriction endonuclease site in the resulting PCR-amplified duplex DNA, and the two restriction sites, when cleaved by the restriction endonuclease, form non-overlapping cohesive termini to facilitate subsequent manipulations.

[0049] Particularly preferred are restriction sites that when cleaved provide overhanging termini adapted for termini-specific modifications such as incorporation of a biotinylated nucleotide (e.g., biotinyl deoxy-UTP) to facilitate subsequent manipulations.

[0050] The above described preferred embodiments in an identifier oligonucleotide are summarized in a specific embodiment shown in Figure 1.

[0051] In Figure 1, a PCR-amplified duplex DNA is shown that is derived from an identifier oligonucleotide described in the Examples. The (Z_n) sequence is illustrated in the brackets as the coding sequence and its complementary strand of the duplex is indicated in the brackets as the anticoding strand. The P1 and P2 sequences are shown in detail with a Sty I restriction endonuclease site defined by the P1 sequence located 5' to the bracket and an Apy I restriction endonuclease site defined by the P2 sequence located 3' to the bracket.

[0052] Step 1 illustrates the cleavage of the PCR-amplified duplex DNA by the enzymes Sty I and Apa I to form a modified identifier sequence with cohesive termini. Step 2 illustrates the specific biotinylation of the anticoding strand at the Sty I site, whereby the incorporation of biotinylated UTP is indicated by a B.

[0053] The presence of non-overlapping cohesive termini after Step 1 in Figure 1 allows the specific and directional cloning of the restriction-digested PCR-amplified fragment into an appropriate vector, such as a sequencing vector. In addition, the Sty I was designed into P1 because the resulting overhang is a substrate for a filling-in reaction with dCTP and biotinyl-dUTP (BTP) using DNA polymerase Klenow fragment. The other restriction site, Apa I, was selected to not provide substrate for the above biotinylation, so that only the anticoding strand can be biotinylated.

[0054] Once biotinylated, the duplex fragment can be bound to immobilized avidin and the duplex can be denatured to release the coding sequence containing the identifier nucleotide sequence, thereby providing purified anticoding strand that is useful as a hybridization reagent for selection of related coding strands as described further herein.

c. Linker Molecules

[0055] A linker molecule in a bifunctional molecule of this invention is represented by B in the above formula A-B-C and can be any molecule that performs the function of operatively linking the polypeptide to the identifier oligonucleotide.

[0056] Preferably, a linker molecule has a means for attaching to a solid support, thereby facilitating synthesis of the bifunctional molecule in the solid phase. In addition, attachment to a solid support provides certain features in practicing the screening methods with a library of bifunctional molecules as described herein. Particularly preferred are linker molecules in which the means for attaching to a solid support is reversible, namely, that the linker can be separated from the solid support.

[0057] A linker molecule can vary in structure and length, and provide at least two features: (1) operative linkage to chemical moiety A, and (2) operative linkage to identifier oligonucleotide C. As the nature of chemical linkages is diverse, any of a variety of chemistries may be utilized to effect the indicated operative linkages to A and to C, as the nature of the linkage is not considered an essential feature of this invention. The size of the linker in terms of the length between A and C can vary widely, but for the purposes of the invention, need not exceed a length sufficient to provide the linkage functions indicated. Thus, a chain length of from at least one to about 20 atoms is preferred.

[0058] Preferred linker molecules are described in Example 3 herein that contains the added, preferred, element of a reversible means for attachment to a solid support. That is, the bifunctional molecule is removable from the solid support after synthesis. In another embodiment, the linker has the ability to be cleaved slowly over time to release the bifunctional molecule, or portions thereof.

[0059] A variety of chemo- or enzyme-selective cleavage functionalities could be incorporated into a linker-support molecule useful in the present invention. For example, a 4-hydroxymethyl phenoxyacetic acid moiety provides an acid-cleavable linker. A 2-[(tert-butyldiphenylsiloxy)methyl]benzioc acid moiety provides a fluoride-cleavable moiety. The phosphate of a 2-hydroxymethyl benzoic acid moiety provides a site cleavable by the combination of alkaline phosphatase treatment followed by mild alkaline treatment. Thus, the incorporation of selectively cleavable linkers other than those recited above are also considered a part of the invention.

[0060] Solid supports for chemical synthesis are generally well known. Particularly preferred are the synthetic resins and controlled pore glass (CPG) supports used in oligonucleotide and in polypeptide synthesis that are available from a variety of commercial sources including Glen Research (Herndon, VA), Bachem Biosciences (Philadelphia, PA), Sigma Chemical Co. (St. Louis, MO), CPG Inc., (Fairfield, NJ) and Applied Biosystems (Foster City, CA). Most preferred are teflon and CPG supports such as are described in Example 2.

[0061] In a related embodiment, the invention describes a preferred bifunctional solid support particularly suited for producing a bifunctional molecule of this invention having a polypeptide as the chemical moiety (i.e., an oligopeptide/oligonucleotide conjugate).

[0062] A preferred bifunctional solid support for synthesizing oligopeptide/oligonucleotide conjugates comprises a solid support, a first linkage unit, a second linkage unit, and a bifunctional unit. A "bifunctional unit" in this context is not

to be confused with a bifunctional molecule of this invention, and refers instead to that chemical moiety present on a bifunctional solid support which provides the two (bi-) reactive functionalities, one for coupling the peptide and another for coupling the oligonucleotide. An exemplary bifunctional unit is shown in Figure 3 as the serine-branch monomer following the aminohexanol linker.

[0063] The solid support used herein is of a type which is dispersible in aqueous solution, such as were described above. A preferred solid support is a controlled pore glass of a type employable for peptide and oligonucleotide synthesis, e.g. aminopropyl-CPG supplied by Sigma. The solid support is tightly coupled to the first linkage unit. A preferred first linkage unit includes a sarcosine linker coupled to the aminopropyl-CPG and a succinyl linker coupled to the sarcosine linker by means of an amide bond. The second linkage unit is, in turn, coupled to the first linkage unit. A preferred second linkage unit is an aminohexanol group. A preferred bond for coupling the first and second linkage units is an alkyl ester. Alkyl esters are readily hydrolysed or cleaved upon exposure to concentrated aqueous ammonia.

[0064] The bifunctional unit has a first leaving group employable for oligopeptide synthesis and a second leaving group employable for oligonucleotide synthesis. A preferred bifunctional unit is an L-serine residue. The L-serine residue includes an amino end, a carboxyl end, and a hydroxyl end. The serine residue is coupled at its carboxyl end to the second linkage unit. In a preferred embodiment, the carboxyl end of the serine residue is couple by means of an amide bond to an aminohexanol linker. The serine residue is also coupled at its amino end to the first leaving group and at its hydroxyl end to the second leaving group. A preferred first leaving group is N-FMOC (N-(9-fluorenylmethoxycarbonyl] or its functional equivalent. A preferred second leaving group is O-DMT (O-dimethoxytrityl) or its functional equivalent.

[0065] The solid support, the first linkage unit, the second linkage unit, the cleavable bond therebetween, and the bifunctional unit, exclusive of said first and second leaving groups, are each substantially chemically unreactive under conditions employed during conventional oligopeptide synthesis protocols using FMOC leaving groups [see: Bodanszky et al., in *The practice of Peptide Synthesis*, Springer-Verlag, (1984); and Bodanszky et al., in *Principles of Peptide Synthesis*, Springer-Verlag, (1984)] and under conditions employed during conventional oligonucleotide synthesis protocols using O-DMT leaving groups and phosphoramidite donors.

[0066] In an alternative embodiment, the bifunctional solid support also includes a third linkage unit. The third linkage unit is interposed between and coupled to the bifunctional unit and the first leaving group. In a preferred mode of this embodiment, the third linkage unit is photosensitive, i.e. it is cleavable by exposure to ultra-violet light. The third linkage unit may include a 3-nitro-4-bromomethyl benzoate group coupled by means of an amide bond to the amino end of the bifunctional serine residue and coupled by means of an ester bond to an FMOC blocked amino acid.

2. Libraries

[0067] A library of this invention is a repertoire of chemical diversity comprising a plurality of species of bifunctional molecules according to the present invention. The plurality of species in a library defines a family of chemical diversity whose species each have a different polypeptide moiety.

[0068] The number of different species in a library represents the complexity of a library and is defined by the length of the polypeptide, and by the size of the amino acid alphabet that can be used to build the polypeptide. The number of different species referred to by the phrase "plurality of species" in a library can be defined by the formula V^a, i.e., V to power of a (exponent a). V represents the alphabet size, i.e., the number of different amino acids X available for use in the polypeptide. "a" is an exponent to V and represents the number of amino acid residues of X forming the polypeptide A, i.e., the length of polypeptide A.

[0069] For example, for a bifunctional molecule where polymer A is a peptide having a length of 6 amino acids, and where the amino acids utilized can be any of the 20 natural amino acids, the alphabet (V) is 20 and the polymer length (a) is 6, and the library size is 20^6 or 64 million. This exemplary library provides a repertoire of chemical diversity comprising 64 million different hexameric polypeptides operatively linked to corresponding unique identifier oligonucleotides.

[0070] Because the complexity of the library will determine the amount of a particular species of bifunctional molecule relative the other species in the library, there are theoretical limits to the maximum useful complexity in a library. Another limitation is the library size defined by the number of actual support molecules or beads present in the library. Therefore it is useful to consider how large (complex) a library should be. This size limit is dictated by the level of sensitivity for detecting the presence of a polymer identifier oligonucleotide after a screening procedure according to this invention. Detection sensitivity is dictated by the threshold of binding or catalytic activity between an acceptor molecule to be assayed and a bifunctional molecule.

[0071] If, for example, the binding threshold is 10^{-6}M (micromolar), then there must be at least one nanomole of each species in a library of 1 milliliter (ml) volume. At this threshold, a library having a complexity of 10^4 could contain 10 micromoles of each species. Because of the reciprocal relationship between library complexity and binding threshold, more complex libraries are possible where the binding threshold is lower.

[0072] The relative amounts of the individual bifunctional molecule species within the library can vary from about

0.2 equivalents to about 10 equivalents, where an equivalent represents the average amount of a species within the library. Preferably each species is present in the library in approximately equimolar amounts.

[0073] In a preferred embodiment, a library contains the complete repertoire of chemical diversity possible based on the mathematical combinations for a given library where there is a fixed alphabet and a preselected number of chemical units in all species of the library. Thus a complete repertoire is one that provides a source of all the possible chemical diversity that can be found in a library of this invention having a fixed alphabet and chemical length.

[0074] It is particularly preferred that a library be comprised of bifunctional molecules where each species of bifunctional molecule contains the same nucleotide sequence for either the P1 or P2 PCR primer binding sites. A library with this design is particularly preferred because, when practicing the methods of this invention, a single PCR primer pair can be used to amplify any species of identifier oligonucleotide (coding sequence) present in the library.

[0075] In a related embodiment, the invention contemplates an element of the library, namely an oligopeptide/oligonucleotide conjugate as described earlier.

[0076] Thus, an element of a library of oligopeptide/oligonucleotide conjugates in one embodiment comprises a solid support, a first linkage unit, a second linkage unit, a bifunctional unit, an oligopeptide, and an oligonucleotide. The solid support is of a type which is dispersible in aqueous solution. The first linkage unit is coupled to the solid support. The second linkage unit is coupled to the first linkage unit. The bifunctional unit is coupled to the second linkage unit. The oligopeptide and the oligonucleotide are attached to the bifunctional unit.

[0077] In an alternative embodiment, the library element includes a cleavable bond for coupling the first linkage unit to the second linkage unit. The cleavable bond is of a type which is cleavable by exposure to concentrated aqueous ammonia.

[0078] In another alternative embodiment, the library element includes a cleavable bond for coupling the bifunctional unit to the oligopeptide. The cleavable bond may be of a type which is cleavable by exposure to ultraviolet light, e.g. a 3-nitro-4-bromomethyl benzoate group coupled by means of an amide bond to the amino end of the bifunctional serine residue and coupled by means of an ester bond to an FMOC blocked amino acid.

[0079] The library element may also be made without the solid support and the first and/or second linkage units. In this instance, the library element comprises a bifunctional unit, an oligopeptide attached to said bifunctional unit, and an oligonucleotide attached to said bifunctional unit. An alternative to this embodiment includes a cleavable bond for coupling the bifunctional unit to the oligopeptide. The cleavable bond may be of a type which is photosensitive, i.e. cleavable by exposure to ultraviolet light, as indicated above.

B. <u>Methods for Producing a Library</u>

[0080] The present method for producing a plurality of bifunctional molecules to form a library of this invention solves a variety of problems regarding efficient synthesis of large numbers of different species.

[0081] In the present synthesis methods, the sequential steps of first adding an amino acid residue X followed by the addition of an oligonucleotide sequence to the linker molecule requires an alternating parallel synthesis procedure to add amino acid residue X and then add a unit identifier nucleotide sequence Z that defines (codes for) that corresponding amino acid residue. The library is built up by the repetition of this alternating parallel process after pooling and division of the reaction products as described herein.

[0082] The only constraint for making an encoded library is that there must be compatible chemistries between the two alternating syntheses procedures for adding an amino acid residue as compared to that for adding an oligonucleotide sequence.

[0083] The problem of synthesis compatibility is solved by the correct choice of compatible protecting groups as the alternating polymers are synthesized, and by the correct choice of methods for deprotection of one growing polymer selectively while the other growing polymer remains blocked, such as by the use of transient protection groups, or the use of protection groups removable under specific deprotection chemistry. Suitable compatible protection chemistries are described herein, and additional suitable chemistries are also described by the chemical protecting group database, available from Synopsys, Inc., or described by Greene et al., in "Protective Groups in Organic Synthesis", 2nd Ed., John Wiley & Sons (1991).

[0084] The synthesis of a library having a plurality of bifunctional molecules comprises the following steps:

(1) A linker molecule is provided that has suitable means for operatively linking the first amino acid residue X_1 and for operatively linking the first nucleotide sequence defining a unit identifier nucleotide Z_1 whose sequence codes for (defines) the structure of X_1. Preferably the linker has a means for attachment to a solid support, and as such allows for the synthesis to proceed in the solid phase.

Thus the provided linker molecule has a structure A'-B-C', where A' represents a termini adapted for reaction to operatively link an amino acid residue X in precursor form (X'), and C' represents a termini adapted for reaction to operatively link an identifier oligonucleotide Z in precursor form (Z'). The termini A' and C' are protected by

respective blocking groups so that during operative linking reactions at one termini, the other termini is protected from reaction.

(2) The linker molecule is then subjected to a first cycle of synthesis to add a building block at one termini. The order of synthesis is not generally important insofar as one may elect to add an amino acid residue X first to termini A', or add an identifier oligonucleotide Z first to termini C'. A first cycle involves the steps of deprotecting the termini of the linker to which a building block is to be added and then adding the building block to the termini. Typically, the added building block contains a blocking group at its free termini, i.e., the termini that will participate in an addition of the next building block of its type. The linker molecule is then subjected to a second cycle of synthesis to add a building block at the other (second) termini. A second cycle involves the steps of deprotecting the second termini of the linker to which a building block is to be added and then adding the building block to the termini. Again, the added building block is typically blocked at its free termini.

The addition of identifier oligonucleotide Z to termini C' can be conducted either nucleotide by nucleotide to form the complete unit identifier nucleotide sequence Z, or Z can be presynthesized, and the oligonucleotide Z added as a block to termini C'. Insofar as the synthesis of oligonucleotides is well known in the arts, the presynthesis of oligonucleotides, and their addition to the growing nucleotide polymer in blocks is preferred because it reduces the number of manipulations in synthesizing a bifunctional molecule.

An amino acid residue X or a unit identifier oligonucleotide Z is referred to as a precursor (X' or Z') to indicate that it contains a leaving group compatible with the reaction chemistry that facilitates the precursor's operative linkage to the growing polymer at the appropriate termini.

The product resulting from step (2) is a bifunctional molecule having the structure $A'\text{-}X_1\text{-}B\text{-}Z_1\text{-}C'$, and is ready for a repetition of the above first and second cycles to add X_2 and Z_2 to the growing polymers.

(3) After the bifunctional molecule product $A'\text{-}X_1\text{-}B\text{-}Z_1\text{-}C'$ is formed, aliquots of the product are made, and the cycles in step (2) are repeated on each aliquot, with the exception that a different species of X (and its corresponding Z) is added in each different aliquot. The reaction product in each aliquot has the structure $A'\text{-}X_2\text{-}X_1\text{-}B\text{-}Z_1\text{-}Z_2\text{-}C'$.

(4) The aliquots each containing the product $A'\text{-}X_2\text{-}X_1\text{-}B\text{-}Z_1\text{-}Z_2\text{-}C'$ are combined (pooled) to form a mixture of different bifunctional molecules, and the mixture is divided into aliquots. The cycles in step (2) are repeated again on each aliquot, with different X and Z building blocks being added to each aliquot to form the bifunctional molecule product $A\text{-}'X_3\text{-}X_2\text{-}X_1\text{-}B\text{-}Z_1\text{-}Z_2\text{-}Z_3\text{-}C'$.

[0085] The process of pooling, aliquoting and adding a next set of building blocks X and Z are then repeated at positions n=4, 5, 6 ... and so on depending on the length of polymers desired. As the cycles are repeated, and the polymers grow in length, the complexity of the resulting library also increases. For each cycle, the polymer length a increases by one and the library complexity therefore increases exponentionally according to the formula V^a. In preferred embodiments, the cycles are repeated from about 1 to 10 times.

[0086] In a related embodiment the provided linker in step (1) is first divided into aliquots, and the cycles of step (2) are conducted on each aliquot adding a different X and corresponding Z to the linker in each different aliquot. The aliquots are then pooled as before, and the cycles of step (2) can be repeated on one or more aliquots.

[0087] Thus the steps of (i) dividing a linker or pool into aliquots, (ii) parallel addition of X and Z to the linker substrate in separate aliquots, and (iii) pooling of the aliquots, can be cycled (repeated) to sequentially add the chemical units and their corresponding unit identifier oligonucleotides to form the library comprising a plurality of bifunctional molecules each having a different chemical polymer operatively linked through the linker to a corresponding identifier oligonucleotide.

[0088] In a preferred embodiment, a method for forming a library of this invention includes the steps for addition of the PCR primer binding sites P1 and P2 to each of the bifunctional molecules in the library.

[0089] The method is substantially the same as above, but includes the addition of a series of nucleotides or a presynthesized P1 oligonucleotide to the linker molecule provided in step (1) prior to the cycles of step (2) that add X and Z. Because all members of the library are to contain the same P1 sequence, P1 is added to the C' termini of linker molecule A'-B-C' prior to dividing the linker into aliquots and subjecting the aliquots to the cycles of step (2) adding X_1 and Z_1. The resulting product has the formula A'-B-P1-C'.

[0090] Thereafter, the product is aliquoted and cycled as before, resulting in the preparation of the product $A'\text{-}(X_n)_a\text{-}B\text{-}P1\text{-}(Z_n)_a\text{-}C'$, where a indicates the presence of a polymer of length "a".

[0091] Next, the pooled admixture containing product $A'\text{-}(X_n)_a\text{-}B\text{-}P1\text{-}(Z_n)_a\text{-}C'$ is subjected to the addition of a series of nucleotides or a presynthesized oligonucleotide P2 at termini C' to form the product $A'\text{-}(X_n)_a\text{-}B\text{-}P1\text{-}(Z_n)_a\text{-}P2\text{-}C'$. Thus all members of the library contain a common sequence P1 and a common sequence P2 from which universal PCR reactions can be conducted, regardless of the species of bifunctional molecule present from the library.

[0092] In a related embodiment, the method for producing a library of this invention utilizes a bf-CPG solid support, as described herein

1. Polypeptide Libraries

[0093] In one preferred embodiment, the invention contemplates a library, and methods of producing the library, where the bifunctional molecule has a polypeptide for polymer A.

[0094] In this embodiment, the compatible chemistries for sequentially adding amino acids and oligonucleotides to the growing polymers has been developed for the synthesis of an amino acid polymer in the direction of carboxy to amino terminus, and alternatively in the direction of amino to carboxy terminus. Chemistries have also been developed for the synthesis of an oligonucleotide polymer in the direction of 3' to 5', and alternatively in the direction of 5' to 3'. In addition, in each of these syntheses it is preferred that the amino acid side chains (R groups) be blocked for certain amino acid residues where the R group provides an otherwise reactive termini during one of the synthesis or deblocking steps.

[0095] Each type of chemistry will be described in detail herein below.

[0096] For any of the syntheses, the reactive side chains of several amino acids must be blocked. Table 1 below lists those of the natural amino acids which have an R group that preferably contains a blocking group. Any compatible protecting (blocking) group may be utilized, and the invention is not to be so limited to any particular blocking group. Also indicated in Table 1 are preferred blocking groups.

TABLE 1

Amino Acid	Blocking Group
Arginine	N-MTr[1], N-PMC[7]
Histidine	N^{π}-Bum[2], SEM[10], FMOC, DNP
Cysteine	S-Trt[3], Acm[11], S-t-butyl
Tryptophan	N^{i}-CHO, none
Tyrosine	O-TBS[4]
Aspartic acid	O-TSE[5], DMB[12]
Glutamic acid	O-TSE[5], DMB
Serine	O-TBS[4]
Threonine	O-TBS[4]
Lysine	N-Bz[6], TFA[8], TEOC[9]
Asparagine	none
Glutamine	none
Glycine	none
Phenylalanine	none
Methionine	none
Alanine	none
Isoleucine	none
Leucine	none
Valine	none
Proline	none

[1] MTr is N^{9}-4-methoxy-2,3,6-trimethylbenzene sulfonyl.
[2] Bum is tert-butoxymethyl.
[3] Trt is triphenylmethyl.
[5] TSE is trimethylsylilethylester.
[4] TBS is tert-butyl-dimethylsilylester.
[6] Bz is benzyl.
[7] PMC is N_G-2,2,5,7,8-pentamethylchromon-6-sulphonyl.
[8] TFA is trifluoroacetyl.
[9] TEOC is β-(trimethyl silyl)ethoxycarbonyl.
[10] SEM is β-(trimethyl silyl)ethoxymethyl.
[11] Acm is acetamidomethyl.
[12] DMB is dimethoxybenzyl.

[0097] Protected amino acids suitable as a blocked precursor for addition to a bifunctional molecule can be obtained from a variety of commercial vendors including Bachem Biosciences Inc. (Philadelphia, PA), Peninsula Labs (CA), and Nova Biochem (CA). In addition, the preparation of protected amino acids is described at least in Example 1.

a. Polypeptide Synthesis

[0098] For synthesis of a polypeptide on the linker substrate in the direction of carboxy to amino terminus, a free amino terminus on the linker is required that can be conveniently blocked and deblocked as needed. A preferred amino terminus blocking group is a fluorenylmethoxycarbonyl group (FMOC).

[0099] FMOC blocked amino termini are deblocked using 20% (v/v) piperidine in dimethylformamide (DMF), or

using 1,8-diazatricyclo[5,4,0]undec-7-ene (DBU) in dichloromethane (DCM) as is well known for polypeptide synthesis. The amino acid units are added in the form of blocked amino acids having FMOC blocked amino termini and a carboxyl terminus blocked with pentafluorophenyl ester (Opfp) or by the activation of the free acid using BOP, HBTU, TBTU or preferebly with pyBOP. BOP is benzotriazole-1-yl-oxy-tris(dimethylamino)-phosphonium hexafluorophosphate (Castro's reagent). HBTU is 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate. TBTU is 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate. pyBOP is bensotriazol-1-yloxytris (dimethylamino phosphonium hexafluorophosphate. The addition reaction requires the blocked amino acid, dimethylformamide (DMF) and hydroxy-benzotriazole (HOBt) as is well known for peptide synthesis. The resulting product contains an added amino acid residue with a FMOC-blocked amino terminus, ready for deblocking addition of a subsequent blocked amino acid as before .

[0100] For synthesis of a polypeptide on the linker substrate in the direction of amino to carboxy terminus, a free carboxy terminus on the linker is required that can be conveniently blocked and deblocked as needed. A preferred carboxy terminus blocking/activating group is the Opfp ester described before. A carboxy terminus on the linker is produced by reacting a linker with a free amino terminus with succinamide in HOBT and a proton catalyst. Thereafter, the terminus can be modified by reaction with pentafluorophenol in dicyclohexylcarbodiimide (DCC) and ethanol acetate to form an Opfp ester at the free carboxy terminus. The Opfp ester is blocked linker terminus is available for addition reaction with a FMOC-, Opfp- blocked amino acid as before, but with the amino acid adding to the linker in the reverse direction. The resulting product contains an added amino acid residue with an Opfp-blocked terminus, ready to repeat the addition with a subsequent blocked amino acid.

b. Oligonucleotide Synthesis

[0101] Oligonucleotides can be synthesized by a variety of chemistries as is well known. An excellent review is "Oligonucleotide Synthesis: A Practical Approach", ed. M.J. Gait, IRL Press, Oxford, (1984). Preferred oligonucleotide synthesis can be carried out on any of a variety of automated DNA synthesizing machines such as are available from Applied Biosystems Inc., (ABI; Foster City, CA). Additional excellent references for oligonucleotide synthesis and DNA synthesis machines and technology include Caruthers, Science, 230:281-285 (1985); and Caruthers, Acc. Chem. Res., 24:278-284 (1991).

[0102] For synthesis of an oligonucleotide on the linker substrate in the direction of 3' to 5', a free hydroxy terminus on the linker is required that can be conveniently blocked and deblocked as needed. A preferred hydroxy terminus blocking group is a dimethoxytrityl ether (DMT). DMT blocked termini are first deblocked, such as by treatment with 3% dichloroacetic acid in dichloromethane (DCM) as is well known for oligonucleotide synthesis, to form a free hydroxy terminus.

[0103] Nucleotides in precursor form for addition to a free hydroxy terminus in the direction of 3' to 5' require a phosphoramidite moiety having an diisopropyl phosphoramidite at the 3' terminus of a nucleotide. In addition, the oxygen of the phosphoramidite is blocked with a cyanoethyl group (CNE), and the 5' terminus is blocked with a DMT ether.

[0104] The addition of a 5' DMT-, 3' CNE- blocked nucleoside phosphoramidite to a free hydroxyl requires tetrazole in acetonitrile followed by (iodine) oxidation and capping of unreacted hydroxyls with acetic anhydride, as is well known for oligonucleotide synthesis. The resulting product contains an added nucleotide residue with a DMT blocked 5' terminus, ready for deblocking and addition of a subsequent blocked nucleotide as before.

[0105] For synthesis of an oligonucleotide on the linker in the direction of 5' to 3', a free hydroxy terminus on the linker is required as before. However, the blocked nucleotide to be added has the blocking chemistries reversed on its 5' and 3' termini to facilitate addition in the opposite orientation.

[0106] A nucleotide with a free 3' hydroxyl and 5' DMT ether is first blocked at the 3' hydroxy terminus by reaction with TBS-Cl in imidazole to form a TBS ester at the 3' terminus. Then the DMT- blocked 5' terminus is deblocked with DCA in DCM as before to form a free 5' hydroxy terminus. The reagent (N,N-diisopropylamino)(cyanoethyl) phosphonamidic chloride is reacted in tetrahydrofuran (THF) with the 5' deblocked nucleotide to form the aminodiisopropyl-, CNE-blocked phosphoramidite group on the 5' terminus. Thereafter the 3' TBS ester is removed with tetrabutylammonium fluoride (TBAF) in DCM to form a nucleotide with the phosphoramidite-blocked 5' terminus and a free 3' hydroxy terminus. Reaction in base with DMT-Cl adds a DMT ether blocking group to the 3' hydroxy terminus.

[0107] The addition of the 3' DMT-, 5' CNE- blocked phosphoramidited nucleotide to a linker substrate having a free hydroxy terminus then proceeds using the previous tetrazole catalyzed reaction, as is well known for oligonucleotide synthesis. The resulting product contains an added nucleotide residue with a DMT-blocked 3' terminus, ready for deblocking with DCA in DCM and the addition of a subsequent blocked nucleotide as before.

[0108] The above demonstrates that the present bifunctional molecules can be synthesized having polypeptide $(X)_a$ in either orientation and having the polymer identifier oligonucleotide $(Z)_a$ in either orientation. Exemplary is the synthesis described herein in detail to form a library of bifunctional molecules having the oligonucleotide attached to linker through its 3' terminus and having the peptide attached to linker through its carboxy terminus.

[0109] In one preferred embodiment, the order of synthesis orients the polypeptide on the linker such that after addition an added amino acid has a free amino terminus, that is, the polymer is assembled in the direction from carboxy to amino terminus. Exemplary chemistry for this synthesis is described in the Examples.

[0110] The addition of oligonucleotides rather than single nucleotides to the growing polymer identifier nucleotide sequence is an alternate embodiment that is preferred because it affords more rapid and modular assembly of the library. Although the previous synthesis discussions involved single nucleotide base units, the same blocking groups and addition chemistries apply where an oligonucleotide is to be added.

[0111] The synthesis of a oligonucleotide having 5' CNE-blocked and 3' DMT-blocked termini or having 3' CNE-blocked and 5' DMT-blocked termini can readily be prepared using the oligonucleotide synthesis methods presently available and described herein.

[0112] After synthesis of a bifunctional molecule, or library of molecules, the blocking groups at termini and at amino acid side chains are removed. Because of the relative lability of termini, it is preferred that the order of deblocking be selected as to preserve the functionalities, particularly the side chain functionalities.

[0113] In the present preferred embodiment for a polypeptide library, the following sequence of deprotecting is preferred when using the teflon support and 5'BCM3 linker described in Examples 2-3:

1) tetrabutyl ammonium fluoride (TBAF) treatment to remove TBS and TMS ethyl ethers;
2) brief 5 minute treatment with trifluoroacetic acid (TFA) treatment sufficient to remove MTr, Bum, PMC and Trt groups, followed by neutralization for 5 minutes with triethylamine;
3) aqueous ammonia treatments to remove Bz and CNE groups; and
4) cleavage of the bifunctional molecule from the solid support using a periodate oxidation.

[0114] Alternatively, the following sequence of deprotection is used on the CPG linker described in Example 3 either with or without the photoactively cleavable linker moiety:

1) TBAF treatment to remove TBS and TMS ethyl ethers; and
2) aqueous ammonia treatment to remove Bz and CNE groups.

[0115] As indicated, after the library has been synthesized, and after the protecting groups have been removed, the bifunctional molecules may be cleaved off of the solid support, and the released bifunctional molecules separated from the solid phase to form a solution comprising a plurality of bifunctional molecules. Alternatively, the library may be maintained in the form of a plurality of bifunctional molecules in the solid phase.

[0116] Although natural amino acids are used in the Examples, the present invention is not to be so limited. The alphabet of possible amino acid residues can be extended to include any molecule that satisfies the basic chemistry defining an amino acid, namely carboxyl and amino termini. Upon polymerization, an amide bond is formed. Thus the possible amino acids can include L-amino acids, D-amino acids, natural amino acids, non-natural amino acids, and derivatives thereof, including pharmaceutically active molecules.

[0117] In addition, there is no basis to limit the polypeptide backbone connecting the termini to the conventional amino acid structure. The amino and carboxylic acid moieties can be on any backbone having any side group substituents, so long as the side groups are properly blocked as described herein. Previously undescribed amino acids may be developed that can be used in the present invention, having heteroatomic configurations, including unusual heterocyclic rings, such as thiazole-alanine or purine alanine.

[0118] The development and use of both conventional and unusual amino acid structures provide a greater diversity of chemical moieties for a library of this invention. Such libraries allow the exploration by the screening methods of this invention of new combinations of important core chemical structures.

[0119] Typical backbones can be alkyl chains of $(CH_2)_n$ where n can be from 1 to at least 6. In addition, the alphabet can comprise amino acids of varying backbone structures such as α, β or γ amino acids. Alphabets can also comprise amino acids where the number of carbon atoms and their configuration in the backbone can be varied.

C. Methods for Identifying Chemical Structures

[0120] The library of this invention provides a repertoire of chemical diversity such that each chemical moiety is linked to a genetic tag that facilitates identification of the chemical structure.

[0121] By the present screening methods, one can identify optimized chemical structures that participate in binding interactions or chemical catalysis events with a biologically active molecule by drawing upon a repertoire of structures randomly formed by the combinatorial association of diverse chemical units without the necessity of either synthesizing them one at a time or knowing their interactions in advance.

[0122] The invention therefore also contemplates a method for identifying a chemical structure that participates in

a preselected binding or catalysis interactions between the chemical structure and a biologically active molecule. The chemical structure to be identified is represented by one of the members of a library of this invention, and the method comprises the following steps:

(1) A library according to the present invention is admixed with a preselected biologically active molecule under binding conditions (i.e., a binding reaction admixture) for a time period sufficient for the biologically active molecule to interact with at least one bifunctional molecule of this invention present in the library and form a binding reaction complex.
(2) The binding reaction complex is then isolated from the library admixture to form an isolated complex.
(3) The nucleotide sequence of the polymer identifier oligonucleotide present in the isolated binding reaction complex is determined. The nucleotide sequence provides a code that defines the chemical structure that participated in the binding reaction, and thus determining that sequence identifies the chemical structure that participates in the binding reaction with the biologically active molecule.

[0123] Where catalysis is the activity to be identified, a reactant can be identified indicating the occurrence of a catalytic event, and the causative solid support in the library is selected as the candidate catalytic molecule.
[0124] A typical biologically active molecule exhibiting a preselected binding interaction or catalytic reaction can be any of a variety of molecules that bind selectively to and/or react with another molecule, including antibodies to antigens, lectins to oligosaccharides, receptors to ligands, enzymes to substrates and the like mediators of molecular interactions and can be catalytic molecules like proteases. Therefore, a preselected binding interaction is defined by the selection of the biologically active molecule with which a library member is to bind. Similarly, a preselected catalytic activity is defined by selection of a substance with which a library member is catalytically active. The binding reaction will be discussed as exemplary, although it is understood that the procedures can be readily adapted for use to detect catalytic polymers.

1. Binding Reaction Admixtures

[0125] The admixture of a library of the invention with a biologically active molecule can be in the form of a heterogeneous or homogeneous admixture. Thus, the members of the library can be in the solid phase with the biologically active molecule present in the liquid phase. Alternatively, the biologically active molecule can be in the solid phase with the members of the library present in the liquid phase. Still further, both the library members and the biologically active molecule can be in the liquid phase.
[0126] Binding conditions are those conditions compatible with the known natural binding function of the biologically active molecule. Those compatible conditions are buffer, pH and temperature conditions that maintain the biological activity of the biologically active molecule, thereby maintaining the ability of the molecule to participate in its preselected binding interaction. Typically, those conditions include an aqueous, physiologic solution of pH and ionic strength normally associated with the biologically active molecule of interest.
[0127] For example, where the binding interaction is to identify a member in the library able to bind an antibody molecule, the preferred binding conditions would be conditions suitable for the antibody to immunoreact with its immunogen, or a known immunoreacting antigen. For a receptor molecule, the binding conditions would be those compatible with measuring receptor-ligand interactions.
[0128] A time period sufficient for the admixture to form a binding reaction complex is typically that length of time required for the biologically active molecule to interact with its normal binding partner under conditions compatible with interaction. Although the time periods can vary depending on the molecule, admixing times are typically for at least a few minutes, and usually not longer than several hours, although nothing is to preclude using longer admixing times for a binding reaction complex to form.
[0129] A binding reaction complex is a stable product of the interaction between a biologically active molecule and a bifunctional molecule of this invention. The product is referred to as a stable product in that the interaction is maintained over sufficient time that the complex can be isolated from the rest of the members of the library without the complex becoming significantly disassociated.

2. Isolation of a Bifunctional Molecule from the Binding Reaction Admixture

[0130] A binding reaction complex is isolated from the binding reaction admixture by any separation means that is selective for the complex, thereby isolating that species of bifunctional molecule which has bound to the biologically active molecule. There are a variety of separation means, depending on the status of the biologically active molecule.
[0131] For example, the biologically active molecule can be provided in admixture in the form of a solid phase reagent, i.e., affixed to a solid support, and thus can readily be separated from the liquid phase, thereby removing the

majority of species of bifunctional molecule. Separation of the solid phase from the binding reaction admixture can optionally be accompanied by washes of the solid support to rinse bifunctional molecules having lower binding affinities off of the solid support.

[0132] Alternatively, for a homogeneous liquid binding reaction admixture, a secondary binding means specific for the biologically active molecule can be utilized to bind the molecule and provide for its separation from the binding reaction admixture.

[0133] For example, an immobilized antibody immunospecific for the biologically active molecule can be provided as a solid phase-affixed antibody to the binding reaction admixture after the binding reaction complex is formed. The immobilized antibody immunoreacts with the biologically active molecule present in the binding reaction admixture to form an antibody-biologically active molecule immunoreaction complex. Thereafter, by separation of the solid phase from the binding reaction admixture, the immunoreaction complex, and therefor any binding reaction complex, is separated from the admixture to form isolated bifunctional molecule.

[0134] Alternatively, a binding means can be operatively linked to the biologically active molecule to facilitate its retrieval from the binding reaction admixture. Exemplary binding means are one of the following high affinity pairs: biotin-avidin, protein A-Fc receptor, ferritin-magnetic beads, and the like. Thus, the biologically active molecule is operatively linked (conjugated) to biotin, protein A, ferritin and the like binding means, and the binding reaction complex is isolated by the use of the corresponding binding partner in the solid phase, e.g., solid-phase avidin, solid-phase Fc receptor, solid phase magnetic beads and the like.

[0135] The use of solid supports on which to operatively link proteinaceous molecules is generally well known in the art. Useful solid support matrices are well known in the art and include cross-linked dextran such as that available under the tradename SEPHADEX from Pharmacia Fine Chemicals (Piscataway, NJ); agarose, borosilicate, polystyrene or latex beads about 1 micron to about 5 millimeters in diameter, polyvinyl chloride, polystyrene, cross-linked polyacrylamide, nitrocellulose or nylon-based webs such as sheets, strips, paddles, plates microtiter plate wells and the like insoluble matrices.

3. Determining the Identifier Sequence

[0136] The nucleotide sequence of the identifier oligonucleotide present in the isolated bifunctional molecules is determined to identify the species of chemical moiety that participated in the preselected binding interaction or catalytic reaction.

[0137] Although conventional DNA sequencing methods are readily available and useful for this determination, the amount and quality of isolated bifunctional molecule may require additional manipulations prior to a sequencing reaction.

[0138] Where the amount is low, it is preferred to increase the amount of the identifier oligonucleotide by polymerase chain reaction (PCR) using PCR primers directed to the primers P1 and P2 present in the identifier oligonucleotide.

[0139] In addition, the quality of the isolated bifunctional molecule may be such that multiple species of bifunctional molecule are co-isolated by virtue of similar capacities for binding to the biologically active molecule. In cases where more than one species of bifunctional molecule are isolated, the different isolated species must be separated prior to sequencing of the identifier oligonucleotide.

[0140] Thus in one embodiment, the different identifier oligonucleotides of the isolated bifunctional molecules are cloned into separate sequencing vectors prior to determining their sequence by DNA sequencing methods. This is typically accomplished by amplifying all of the different identifier oligonucleotides by PCR as described herein, and then using the unique restriction endonuclease sites on the amplified product as shown in Figure 1 to directionally clone the amplified fragments into sequencing vectors. The cloning and sequencing of the amplified fragments then is a routine procedure that can be carried out by any of a number of molecular biological methods known in the art.

[0141] Alternatively, PCR amplified products derived from a population of isolated bifunctional molecules can be used as a hybridization probe to selectively enrich the quality of the isolated bifunctional molecules. For example, using the hybridization probes, which are modified by biotinylation as shown in Figure 1, one can isolate members of the library by hybridization, to form an enriched library containing only bifunctional molecules that have sequences that hybridize to the above hybridization probes. In a second screening reaction under different binding conditions, for example, higher stringency binding conditions, one can isolate the species of bifunctional molecule that binds most tightly with the biologically active molecule.

[0142] Thus the library can be manipulated to form enriched libraries from which to screen for chemical diversity.

4. Polymerase Chain Reaction

[0143] For determining the nucleotide sequence of the identifier oligonucleotide in the isolated complex as part of the methods of this invention, the use of the polymerase chain reaction (PCR) is a preferred embodiment.

[0144] For use in this invention, the identifier oligonucleotide are comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the genetic material to be assayed is in the form of double stranded DNA, it is usually first denatured, typically by melting, into single strands. The nucleic acid is subjected to a PCR reaction by treating (contacting) the sample with a PCR primer pair, each member of the pair having a preselected nucleotide sequence. The PCR primer pair is capable of initiating primer extension reactions by hybridizing to the PCR primer binding site on nucleotide sequences of the identifier nucleotide, preferably at least about 10 nucleotides in length, and more preferably is at least about 20 nucleotides in length. The first primer of a PCR primer pair is sometimes referred to herein as the "anti-sense primer" because it hybridizes to a non-coding or anti-sense strand of a nucleic acid, i.e., a strand complementary to a coding strand. The second primer of a PCR primer pair is sometimes referred to herein as the "sense primer" because it hybridizes to the coding or sense strand of a nucleic acid.

[0145] The PCR reaction is performed by mixing the PCR primer pair, preferably a predetermined amount thereof, with the nucleic acids of the sample, preferably a predetermined amount thereof, in a PCR buffer to form a PCR reaction admixture. The admixture is thermocycled for a number of cycles, which is typically predetermined, sufficient for the formation of a PCR reaction product, thereby enriching the sample to be assayed for the identifier oligonucleotide in the isolated complex.

[0146] PCR is typically carried out by thermocycling i.e., repeatedly increasing and decreasing the temperature of a PCR reaction admixture within a temperature range whose lower limit is about 30 degrees Celsius (30°C) to about 55°C and whose upper limit is about 90°C to about 100°C. The increasing and decreasing can be continuous, but is preferably phasic with time periods of relative temperature stability at each of temperatures favoring polynucleotide synthesis, denaturation and hybridization.

[0147] A plurality of first primer and/or a plurality of second primers can be used in each amplification, e.g., one species of first primer can be paired with a number of different second primers to form several different primer pairs. Alternatively, an individual pair of first and second primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and second primers can be combined for assaying for mutations.

[0148] The PCR reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, i.e., a PCR buffer, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

[0149] The PCR buffer also contains the deoxyribonucleotide triphosphates (polynucleotide synthesis substrates) dATP, dCTP, dGTP, and dTTP and a polymerase, typically thermostable, all in adequate amounts for primer extension (polynucleotide synthesis) reaction. The resulting solution (PCR admixture) is heated to about 90°C - 100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to 54°C, which is preferable for primer hybridization. The synthesis reaction may occur at from room temperature up to a temperature above which the polymerase (inducing agent) no longer functions efficiently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature is generally no greater than about 40°C. The thermocycling is repeated until the desired amount of PCR product is produced. An exemplary PCR buffer comprises the following: 50 mM KCl; 10 mM Tris-HCl at pH 8.3; 1.5 mM $MgCl_2$; 0.001% (wt/vol) gelatin, 200 μM dATP; 200 μM dTTP; 200 μM dCTP; 200 μM dGTP; and 2.5 units Thermus aquaticus (Taq) DNA polymerase I (U.S. Patent No. 4,889,818) per 100 microliters (μl) of buffer.

[0150] The inducing agent may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be inducing agents, however, which initiate synthesis at the 5' end and proceed in the above direction, using the same process as described above.

[0151] The inducing agent also may be a compound or system which will function to accomplish the synthesis of RNA primer extension products, including enzymes. In preferred embodiments, the inducing agent may be a DNA-dependent RNA polymerase such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase. These polymerases produce a complementary RNA polynucleotide. The high turn-over rate of the RNA polymerase amplifies the starting polynucleotide as has been described by Chamberlin et al., The Enzymes, ed. P. Boyer, pp. 87-108, Academic Press, New York (1982). Amplification systems based on transcription have been described by Gingeras et al., in PCR Protocols, A Guide to Methods and Applications, pp. 245-252, Innis et al., eds, Academic Press, Inc., San Diego, CA (1990).

[0152] If the inducing agent is a DNA-dependent RNA polymerase and, therefore incorporates ribonucleotide triphosphates, sufficient amounts of ATP, CTP, GTP and UTP are admixed to the primer extension reaction admixture and

the resulting solution is treated as described above.

[0153] The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which can be used in the succeeding steps of the process.

[0154] PCR amplification methods are described in detail in U.S. Patent Nos. 4,683,192, 4,683,202, 4,800,159, and 4,965,188, and at least in several texts including PCR Technology: Principles and Applications for DNA Amplification, H. Erlich, ed., Stockton Press, New York (1989); and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds., Academic Press, San Diego, California (1990).

[0155] The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

[0156] The term "primer" as used herein refers to a polynucleotide whether purified from a nucleic acid restriction digest or produced synthetically, which is capable of acting as a point of initiation of nucleic acid synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase, reverse transcriptase and the like, and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency, but may alternatively be in double stranded form. If double stranded, the primer is first treated to separate it from its complementary strand before being used to prepare extension products. Preferably, the primer is a polydeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agents for polymerization. The exact lengths of the primers will depend on many factors, including temperature and the source of primer. For example, depending on the complexity of the target sequence, a polynucleotide primer typically contains 15 to 25 or more nucleotides, although it can contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template.

[0157] The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be synthesized or amplified. This means that the primer must be sufficiently complementary to non-randomly hybridize with its respective template strand. Therefore, the primer sequence may or may not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such non-complementary fragments typically code for an endonuclease restriction site. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarity with the sequence of the strand to be synthesized or amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions.

[0158] Primers of the present invention may also contain a DNA-dependent RNA polymerase promoter sequence or its complement. See for example, Krieg et al., Nucl. Acids Res., 12:7057-70 (1984); Studier et al., J. Mol. Biol., 189:113-130 (1986); and Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

[0159] When a primer containing a DNA-dependent RNA polymerase promoter is used, the primer is hybridized to the polynucleotide strand to be amplified and the second polynucleotide strand of the DNA-dependent RNA polymerase promoter is completed using an inducing agent such as E. coli DNA polymerase I, or the Klenow fragment of E. coli DNA polymerase. The starting polynucleotide is amplified by alternating between the production of an RNA polynucleotide and DNA polynucleotide.

[0160] Primers may also contain a template sequence or replication initiation site for a RNA-directed RNA polymerase. Typical RNA-directed RNA polymerase include the QB replicase described by Lizardi et al., Biotechnology, 6:1197-1202 (1988). RNA-directed polymerases produce large numbers of RNA strands from a small number of template RNA strands that contain a template sequence or replication initiation site. These polymerases typically give a one million-fold amplification of the template strand as has been described by Kramer et al., J. Mol. Biol., 89:719-736 (1974).

[0161] The polynucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester or phosphodiester methods see Narang et al., Meth. Enzymol., 68:90, (1979); U.S. Patents No. 4,356,270, No. 4,458,066, No. 4,416,988, No. 4,293,652; and Brown et al., Meth. Enzymol., 68:109, (1979).

[0162] If the nucleic acid sample is to be enriched for the identifier oligonucleotide in the isolated complex by PCR amplification, two primers, i.e., a PCR primer pair, must be used for each coding strand of nucleic acid to be amplified. The first primer becomes part of the non-coding (anti-sense or minus or complementary) strand and hybridizes to a nucleotide sequence on the plus or coding strand. Second primers become part of the coding (sense or plus) strand and hybridize to a nucleotide sequence on the minus or non-coding strand. One or both of the first and second primers can contain a nucleotide sequence defining an endonuclease recognition site as described herein. The site can be heterologous to the polymer identifier oligonucleotide being amplified.

[0163] In one embodiment, the present invention utilizes a set of polynucleotides that form primers having a priming region located at the 3'-terminus of the primer. The priming region is typically the 3'-most (3'-terminal) 15 to 30 nucle-

otide bases. The 3'-terminal priming portion of each primer is capable of acting as a primer to catalyze nucleic acid synthesis, i.e., initiate a primer extension reaction off its 3' terminus. One or both of the primers can additionally contain a 5'-terminal (5'-most) non-priming portion, i.e., a region that does not participate in hybridization to the preferred template.

5. Nucleic Acid Sequence Analysis

[0164] Nucleic acid sequence analysis is a well known procedure for determining the sequence of nucleotides and is applied to the present methods to determine the nucleotide sequence in an identifier oligonucleotide or PCR amplification product of this invention. Nucleic acid sequence analysis is approached by a combination of (a) physiochemical techniques, based on the hybridization or denaturation of a probe strand plus its complementary target, and (b) enzymatic reactions with endonucleases, ligases; and polymerases.

[0165] In assays using nucleic acid hybridization, detecting the presence of a DNA duplex in a process of the present invention can be accomplished by a variety of means.

[0166] In one approach for detecting the presence of a DNA duplex, an oligonucleotide that is hybridized in the DNA duplex includes a label or indicating group that will render the duplex detectable. Typically such labels include radioactive atoms, chemically modified nucleotide bases, and the like.

[0167] The oligonucleotide can be labeled, i.e., operatively linked to an indicating means or group, and used to detect the presence of a specific nucleotide sequence in a target template.

[0168] Radioactive elements operatively linked to or present as part of an oligonucleotide probe (labeled oligonucleotide) provide a useful means to facilitate the detection of a DNA duplex. A typical radioactive element is one that produces beta ray emissions. Elements that emit beta rays, such as ^{3}H, ^{14}C, ^{32}P and ^{35}S represent a class of beta ray emission-producing radioactive element labels. A radioactive polynucleotide probe is typically prepared by enzymatic incorporation of radioactively labeled nucleotides into a nucleic acid using DNA kinase.

[0169] Alternatives to radioactively labeled oligonucleotides are oligonucleotides that are chemically modified to contain metal complexing agents, biotin-containing groups, fluorescent compounds, and the like.

[0170] One useful metal complexing agent is a lanthanide chelate formed by a lanthanide and an aromatic beta-diketone, the lanthanide being bound to the nucleic acid or oligonucleotide via a chelate-forming compound such as an EDTA-analogue so that a fluorescent lanthanide complex is formed. See U.S. Patent No. 4,374,120, No. 4,569,790 and published Patent Application Nos. EP0139675 and WO87/02708.

[0171] Biotin or acridine ester-labeled oligonucleotides and their use to label polynucleotides have been described. See U.S. Patent No. 4,707,404, published Patent Application EP0212951 and European Patent No. 0087636. Useful fluorescent marker compounds include fluorescein, rhodamine, Texas Red, NBD and the like.

[0172] A labeled oligonucleotide present in a DNA duplex renders the duplex itself labeled and therefore distinguishable over other nucleic acids present in a sample to be assayed. Detecting the presence of the label in the duplex and thereby the presence of the duplex, typically involves separating the DNA duplex from any labeled oligonucleotide probe that is not hybridized to a DNA duplex.

[0173] Techniques for the separation of single stranded oligonucleotide, such as non-hybridized labeled oligonucleotide probe, from DNA duplex are well known, and typically involve the separation of single stranded from double stranded nucleic acids on the basis of their chemical properties. More often separation techniques involve the use of a heterogeneous hybridization format in which the non-hybridized probe is separated, typically by washing, from the DNA duplex that is bound to an insoluble matrix. Exemplary is the Southern blot technique, in which the matrix is a nitrocellulose sheet and the label is ^{32}P. Southern, J. Mol. Biol., 98:503 (1975).

[0174] The oligonucleotides can also be advantageously linked, typically at or near their 5'-terminus, to a solid matrix, i.e., aqueous insoluble solid support as previously described.

[0175] It is also possible to add "linking" nucleotides to the 5' or 3' end of the member oligonucleotide, and use the linking oligonucleotide to operatively link the member to the solid support.

[0176] In nucleotide hybridizing assays, the hybridization reaction mixture is maintained in the contemplated method under hybridizing conditions for a time period sufficient for the oligonucleotides having complementarity to the predetermined sequence on the template to hybridize to complementary nucleic acid sequences present in the template to form a hybridization product, i.e., a complex containing oligonucleotide and target nucleic acid.

[0177] The phrase "hybridizing conditions" and its grammatical equivalents, when used with a maintenance time period, indicates subjecting the hybridization reaction admixture, in the context of the concentrations of reactants and accompanying reagents in the admixture, to time, temperature and pH conditions sufficient to allow one or more oligonucleotides to anneal with the target sequence, to form a nucleic acid duplex. Such time, temperature and pH conditions required to accomplish hybridization depend, as is well known in the art, on the length of the oligonucleotide to be hybridized, the degree of complementarity between the oligonucleotide and the target, the guanine and cytosine content of the oligonucleotide, the stringency of hybridization desired, and the presence of salts or additional reagents in

the hybridization reaction admixture as may affect the kinetics of hybridization. Methods for optimizing hybridization conditions for a given hybridization reaction admixture are well known in the art.

[0178] Typical hybridizing conditions include the use of solutions buffered to pH values between 4 and 9, and are carried out at temperatures from 4°C to 37°C, preferably about 12°C to about 30°C, more preferably about 22°C, and for time periods from 0.5 seconds to 24 hours, preferably 2 minutes (min) to 1 hour. Exemplary are the conditions described in Example 4.

[0179] Hybridization can be carried out in a homogeneous or heterogeneous format as is well known. The homogeneous hybridization reaction occurs entirely in solution, in which both the oligonucleotide and the nucleic acid sequences to be hybridized (target) are present in soluble forms in solution. A heterogeneous reaction involves the use of a matrix that is insoluble in the reaction medium to which either the oligonucleotide, polynucleotide probe or target nucleic acid is bound.

[0180] Where the nucleic acid containing a target sequence is in a double stranded (ds) form, it is preferred to first denature the dsDNA, as by heating or alkali treatment, prior to conducting the hybridization reaction. The denaturation of the dsDNA can be carried out prior to admixture with a oligonucleotide to be hybridized, or can be carried out after the admixture of the dsDNA with the oligonucleotide.

[0181] Effective amounts of the oligonucleotide present in the hybridization reaction admixture are generally well known and are typically expressed in terms of molar ratios between the oligonucleotide to be hybridized and the template. Preferred ratios are hybridization reaction mixtures containing equimolar amounts of the target sequence and the oligonucleotide. As is well known, deviations from equal molarity will produce hybridization reaction products, although at lower efficiency. Thus, although ratios where one component can be in as much as 100 fold molar excess relative to the other component, excesses of less than 50 fold, preferably less than 10 fold, and more preferably less than two fold are desirable in practicing the invention.

Examples

[0182] The following examples are intended to illustrate, but not limit, the present invention.

1. Preparation of Protected Amino Acids

[0183] The synthesis of a bifunctional molecule requires protected amino acids. The amino-terminus of the amino acid is protected with fluorenylmethoxycarbonyl (FMOC) and the carboxy-terminus is protected with a pentafluorophenyl ester (Opfp). The amino acids lysine, cysteine, tyrosine, serine, threonine, arginine, histidine, tryptophan, aspartic acid, and glutamic acid, require additional protection of their side chains (R groups).

[0184] Most of the FMOC and Opfp protected amino acids are commercially available and were obtained from Bachem Biosciences, Inc. (Philadelphia, PA). The terminology used herein for their structure is indicated by the following example for glycine (Gly): FMOC-Gly-Opfp, where FMOC and Opfp are the amino and carboxy terminal protecting groups. For side chain protection, the following protected amino acids are available from Bachem: FMOC-Arg (MTr)-Opfp having the substituent N^{9}-4-methoxy-2,3,6-trimethylbenzene sulfonyl arginine (MTr) at the side chain amino terminus of arginine; FMOC-His(Bum)-Opfp having the substituent N^{π}-tert-butoxymethylhistidine (Bum) at the heterocyclic reactive nitrogen in histidine; FMOC-Cys(Trt)-Opfp having the substituent S-triphenyl methyl cysteine at the side chain sulfur of cysteine; FMOC-Trp(N-For)-Opfp having a formyl group at the amino group of tryptophan's heterocyclic group; and FMOC-Lys(N-Bz)-Opfp having a benzyl group on the free amino group of lysine's side chain; where the structure in parenthesis indicates the protecting group on the reactive side chain.

[0185] FMOC-Tyr(OTBS)-Opfp having a tertbutyldimethysilyl(TBS)ester on the side chain hydroxy of tyrosine is prepared by reacting an excess of formic acid with FMOC-Tyr(tert-butyl)-Opfp (Bachem) to remove the tert butyl group from the protected hydroxyl group to form FMOC-Tyr-Opfp. Thereafter, one equivalent of FMOC-Tyr-Opfp is reacted with 1.2 equivalents of TBS-Cl and 1.5 equivalents of imidazole in DCM at room temperature for 12 hours under inert atmosphere to form FMOC-Tyr(OTBS)-Opfp.

[0186] FMOC-Ser(OTBS)-Opfp is similarly prepared using FMOC-Ser(tert-butyl)-Opfp (Bachem) in the reaction. FMOC-Thr(OTBS)-Opfp is also prepared in this manner using FMOC-Thr(tert-butyl)-Opfp (Bachem).

[0187] FMOC-Asp(TMSE)-Opfp having a trimethylsilyl ethyl ester (TMSE) on the side chain carboxyl group of aspartic acid is prepared by first reacting one equivalent of FMOC-Asp-O-tertbutyl (Bachem) with 1.5 equivalents of 2-trimethylsilylethanol and 1.5 equivalents if dicyclocarbodiimide (DCC) in ethyl acetate for 12 hours at room temperature under inert atmosphere to form FMOC-Asp(OTMSE)-O-tertbutyl. There after the TMSE ester is reacted with an excess of formic acid at room temperature for 14 hours to hydrolyze the tertbutyl moiety and form a free carboxyl terminus in the form of FMOC-Asp(OTMSE)-COOH. The formic acid is evaporated, and the 1 equivalent of the remaining amino acid is admixed with 1.1 equivalent of pentafluorophenol (pfp; Bachem) and 1.1 equivalent of DCC for 12 hours at room temperature under inert atmosphere to form the product FMOC-Asp(TMSE)-Opfp. The product is isolated from unre-

acted pfp, DCC and precursor amino acid by silica gel chromatography using 10% (v/v) ethyl acetate in hexane.

[0188] FMOC-Glu(TMSE)-Opfp having a TMSE ester on the side chain carboxyl group of glutamic acid is prepared as described above to prepare FMOC-,TMSE- and pfp protected aspartic acid, except that FMOC-Glu-O-tert-butyl (Bachem) is used in place of the aspartic acid precursor, to form FMOC-Glu(TMSE)-Opfp.

2. Preparation of Solid Support

a. Teflon Support

[0189] A solid support designated N6-(5'-O-dimethyoxytrityl-2',3'-diacetyl 1-adenylyl)-teflon support was obtained from Glen Research (Herndon, VA). The solid support is a teflon resin with a modified adenine nucleoside having the solid support linkage through the 6-amino group of a purine base, a dimethoxytrityl ether (DMT) at the 5' position of the ribose ring and acetate esters at the 2' and 3' positions of the ribose ring. The solid support was admixed with 5 volumes of 3% (v/v) dichloroacetic acid in dichloromethane (3% DCA in DCM) and maintained for 10 minutes at room temperature under inert atmosphere to remove the dimethoxytrityl (DMT) protecting group to form a free 5'hydroxyl. The resulting deprotected solid support was washed 3 times with DCM to remove the excess unreacted DCA. The deprotected washed teflon solid support is ready for coupling to linker.

3. Coupling of Linker to Solid Support

a. Teflon Support

[0190] One equivalent of deprotected washed solid support (prepared in Example 2) was admixed with 20 equivalents of a linker designated (1-dimethoxytrityloxy-3-fluorenyl methoxycarbonylaminopropan-2yl)-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite and referred to as 5' Branched-Modifier C3 (or 5'BMC3; available from Glen Research) and 20 equivalents of tetrazole (0.45M in acetonitrile, Glen Research) and was agitated with stirring for 0.5 to 1 hour at room temperature to under inert atmosphere. The admixture was then washed with an excess of acetonitrile to remove unreacted reagents. The washed solid phase material was retained and admixed with 2 equivalents of Iodine in tetrahydrofuran/water, 9:1, (available from Glen Research) under inert atmosphere and maintained at room temperature for 10 minutes to form oxidized solid-support coupled linker.

[0191] One equivalent of oxidized solid support-coupled linker was then admixed with 20 equivalents of acetonitrile/acetic anhydride, 88:12, (capping reagent; Glen Research) for 10 minutes at room temperature to cap any unreacted free hydroxyls present on the solid support and form capped teflon-solid support-coupled linker.

b. Controlled Pore Glass (CPG) Support

[0192] A controlled pore glass (CPG) support and linker is particularly preferred that can also be used in the present invention. The complete structure of an exemplary CPG support is shown in Figure 3, and includes (1) a sarcosine linker moiety connecting the linker to the CPG support, (2) a succinyl-aminohexanol linker that is cleavable by aqueous ammonia to release the polypeptide-linker-oligonucleotide conjugate (A-B-C) from the solid support, and (3) a serine branch-monomer having a photolabile cleavage site for releasing the oligonucleotide from the solid support. The synthesis is carried out stepwise as described below.

[0193] **N-Fmoc-amidohexan-1-ol (1)** was first prepared to form the aminohexanol linker moiety. To that end, 6-Amino-1-hexanol (0.75 g, 6.4 mmol) was dissolved in sat. aq. Na_2CO_3 (10 ml) and cooled on ice. 9-Fluorenylmethyl chloroformate (Fmoc-Cl, 1.83 g, 7.1 mmol) in THF (25 ml, freshly distilled) was added slowly under vigorous stirring. The solution is acidified with 10% (w/v) citric acid, extracted with ethyl acetate and filtered through Celite 545 before separation of the phases. The organic phase was dried (Na_2SO_4), filtered, and evaporated to yield the crude product (2.15 g). Pure product was obtaind by recrystalization from ethylacetate/hexane (100 ml, 8/2, v/v) to form the alcohol 1. The yield of alcohol 1 was 1.6 g.

[0194] **N-Fmoc-amidohexan-1-yl succinate (2)** was prepared to form the succinyl-aminohexanol linker moiety. The alcohol 1 (758 mg, 2.2 mmol) was dissolved in anh. pyridine (5 ml) and evaporated to dryness *in vacuo*. The evaporated oil was redissolved in anh. pyridine (3 ml) and succinic anhydride (229 mg, 2.3 mmol), 4-dimethylaminopyridine (DMAP, 12.5 mg, 0.1 mmol), and diisopropylethylamine (DIPEA, 370 µl, 278 mg, 2.2 mmol) was added with stirring under an inert atmosphere (Ar). The reaction mixture was stirred over night, evaporated to dryness, redissolved in dichloromethane (100 ml) and extracted with dilute aq. hydrochloric acid (1 M, 50 ml). The organic phase was dried (Na_2SO_4), filtered, and evaporated to yield the crude product as an oil designated compound 2 which could be recrystalized from 2-propanol. The yield of compound 2 was 735 mg (as white crystals), having a melting point (Mp.) of 67-68°C (uncorr.); and a mass spectroscopy (MS) (EI, *m/e*) of 440 (calcd for $C_{25}H_{29}NO_6$ + H 440).

[0195] O-(4,4'-Dimethoxytrityl)-N-Fmoc-*L*-Serine (3) was prepared to form the serine-branch linker moiety containing the FMOC and O-DMT termini. Fmoc-*L*-Serine (4.04 g, 12.2 mmol) was dissolved in anh. pyridine (10 ml) and evaporated to dryness *in vacuo*. This process was repeated twice. The evaporated oil was redissolved in anh. pyridine (12 ml) and 4,4'-dimethoxytrityl chloride (4.20 g, 12.4 mmol) was added with stirring at room temperature under an inert atmosphere (Ar). The reaction mixture was stirred over night, evaporated to dryness, redissolved in chloroform (4 x 25 ml) and extracted with sat. aq. $NaHCO_3$ (50 ml). The aq. phase was back-extracted once with chloroform (25 ml) and the combined organic phases were dried from Na_2SO_4, filtered, and evaporated to yield the crude product as a brownish oil. The oil was dissolved in chloroform/ethylacetate and triturated once with hexane and then purified by silica gel column chromatography (150 ml silica) using $CHCl_3$/MeOH/DIPEA, 94/5/1, as the mobile phase. The fractions were analyzed by tlc and the appropriate fractions were pooled and evaporated to give a white foam designated compound 3. The yield of compound 3 was 4.29 g, and the
MS (EI, *m/e*) was 630 (calcd for $C_{39}H_{35}NO_7$ + H 630).

[0196] Controlled pore glass (CPG) was activated for coupling to the above prepared linkers. To that end, CPG (Sigma G5019, aminopropyl-CPG, 4.97 g) was suspended in trichloroacetic acid in dichloromethane (3%, 20 ml) and agitated in a shaker for 4 hours. The CPG was isolated by filtration, washed three times with dichloromethane (20 ml), three times with DIPEA in chloroform (10%, 20 ml) and three times with diethyl ether to form activated CPG. The activated CPG was dried *in vacuo*.

[0197] Bi-functional CPG [bf-CPG, CPG-Sar-suc-aho-Ser(*O*-DMT)(*N*-Fmoc)] was then prepared comprising all of the element shown in Figure 3.

i. Activated CPG (2.0 g, loading 83 µmole/g) was placed in a filter funnel (25 ml). Fmoc-Sar (313.5 mg, mmol) was dissolved in DMF (2.8 ml), mixed with pyBOP (525.9 mg, mmol) in DMF (2.8 ml) and the combined, activated amino acid mixture was added to the CPG. DIPEA (1 ml) was added to the reaction mixture and the CPG shaked for 1 hour at room temperature. Chloroform (5 ml) was added and the shaking continued for another 3 hours. The Sar-functionalized CPG was recovered by filtration and washed with DMF (2 x), dichloromethane (2 x), diethyl ether (2x) and dried *in vacuo*. To obtain maximal loading, this process was repeated to give CPG-Sar with a loading of 77 µmole/g. Eventual remaining free amino groups were capped with acetic anhydride/DMAP followed by an extensive washing and drying *in vacuo* to form Sar-functionalized CPG.

ii. The Sar-functionalized CPG (2.0 g) was treated with piperidine in DMF (2/8, 2 x 5 minutes) and reacted with compound 2 (450 mg), pyBOP (575 mg) and DIPEA (1000 µl) in DMF (2 x 2.8 ml). The resulting Sar-suc-aho-Fmoc-functionalized CPG was recovered by filtration and washed with DMF (2 x), dichloromethane (2 x), diethyl ether (2x) and dried *in vacuo*.

iii. The Sar-suc-aho-Fmoc-functionalized CPG was treated with piperidine in DMF (2/8, 2 x 10 minutes) and reacted with compound 3 (764 mg), pyBOP (536 mg) and DIPEA (1000 µl) in DMF (2 x 2.8 ml) to form a bifunctional CPG support or "bf-CPG". The loading was determined to be 61.7 µmole/g as judged from the Fmoc/piperidine adduct absorption at 302 nm.

[0198] In an additional embodiment, the above bf-CPG is further functionalized by adding a photolabile cleavage site for removal of the polypeptide from the conjugate.

[0199] To that end, 3-nitro-4-bromomethylbenzoic acid is prepared as described by Rich et al., J. Am. Chem. Soc., 97:1575-1579 (1975) using α-bromo-p-toluic acid (4-bromomethulbenzoic acid; commercially available from Aldrich Chemical Company, Milwaukee, WI). The 3-nitro-4-bromomethybenzoic acid moiety is added to the above bf-CPG support by first deblocking the N-terminal function by a treatment with piperidine in DMF (2/8, 2 x 5 minutes), then washed thoroughly and reacted with the 3-nitro-4-bromomethylbenzoic acid that is activated with pyBOP to form a photocleavable bf-CPG having the structure shown in Figure 3. The completeness of the coupling reaction is monitored by a positive and a negative Kaiser test after the FMOC deprotection and the coupling reaction, respectively.

[0200] Coupling of an amino acid residue Monomer to the phocleavable bf-CPG (also referred to as *o*-NB-bf-CPG) is carried out by the esterification procedure described by Rich et al., supra, and by Barany et al., J. Am. Chem. Soc., 107:4936-4942 (1985)

[0201] The resulting bf-CPG linker/support molecules have novel and useful features in the context of the present invention described hereinbelow.

(1) The linker/support molecule has an aqueous ammonia sensitive cleavage site in the sarcosine/succinyl moiety which is stable to Fmoc peptide synthesis conditions, and which upon cleavage releases the peptide-oligonucleotide conjugate free from the solid CPG support, thereby allowing for the production of a soluble phase form of the chemical library. A soluble phase chemical library is particularly suited for use in PCR reactions, and for measuring peptide interactions in solution phase binding interactions.

(2) The L-serine branch-monomer renders the peptide-oligonucleotide conjugate isomerically pure.

(3) The photolabile linker/support molecule (*o*-NB-bf-CPG) has a light sensitive (hv) cleavage site, indicated by the arrow in Figure 3, that is sensitive to light of about 350 nm and which upon cleavage results in the release of soluble polypeptide. Cleavage can be regulated by the time and intensity of irradiation using, for example, a Rayonet RPR Reactor as described by Barany et al., supra.

4. Coupling of Nucleotide to Solid Support

a. Deprotect DMT

[0202] The capped teflon solid support-coupled linker prepared in Example 3 was admixed with 3% DCA in DCM for 10 minutes at room temperature under inert atmospheres to remove the DMT protecting group from the linker and form a free hydroxyl group. The deprotected linker/support was then washed 3 times with DCM. The deprotected linker/support is ready for addition of nucleotide.

b. Addition of Nucleotide to Teflon Support

[0203] One equivalent of deprotected linker/support was admixed with about 20 equivalents of a desired blocked nucleotide phosphoramidite and 20 equivalents of tetrazole (0.45M in acetonitrile) to form a coupled nucleotide/linker/support (coupled nucleotide complex). The coupled nucleotide complex was then washed with an excess of acetonitrile to remove unreacted reagents. All blocked nucleotide phosphoramidites were obtained from Glen Research and contain a DMT protected blocked 5'hydroxyl, a cyanoethyl ester (CNE) and a diisopropylamine group at the 3'-phosphoramidite. In addition, the adenine and cytosine derivatives contained a benzoyl group on the base's free nitrogen and the guanosine derivative contains an isobutyl group on the 2-amino group of the purine base.

[0204] One equivalent of coupled nucleotide complex was then admixed with 2 equivalents of Iodine in tetrahydrofuran/water, 9:1, for oxidation as before in Example 3 to oxidize the coupled nucleotide complex.

[0205] Thereafter, one equivalent of oxidized nucleotide complex was admixed with 20 equivalents of capping reagent as before in Example 3 to cap any unreacted free hydroxyls and to form capped solid phase-coupled nucleotide complex.

c. Coupling of Nucleotide to Bifunctional CPG

[0206] As a demonstration of coupling chemistry for addition of a nucleotide residue to the bifunctional CPG support, we have synthesized and coupled an oligodeoxyribonucleotide onto bf-CPG. To that end, oligonucleotides were synthesized on a ABI 394 DNA synthesizer using the standard 1 µmole scale synthesis cycle and commercially available reagents and phosphoramidites. The solid support used was the bf-CPG (20 mg, 1 µmol) placed in a commercially available empty synthesis column for the ABI synthesizers. Repetitive yield were calculated to 98.9% as judged from collecting the detritylation, deluting in *p*-toluene sulfonic acid monohydrate in acetonitril (0.1 M) and measuring the absorptions at 498 nm.

[0207] The oligonucleotide was released from the support by conc. aq. ammonia (shaking for more than 24 hours). The oligonucleotide was isolated using commercially available OPC-cartridges and analyzed by HPLC and PAGE (radiolabeled with ^{32}P-g-ATP and T4-kinase).

5. Coupling of Amino Acid to Solid Support

a. Deprotect Nucleotide Complex

[0208] One equivalent of capped teflon solid phase-coupled nucleotide complex was admixed with 1 equivalent of 1,8,diazabicyclo[5,4,0]undec-7-ene (DBU; available from Aldrich Chemical Co., Milwaukee, WI) in DCM under inert atmosphere for 10 minutes at room temperature to remove (deblock) the fluoromethoxycarbonyl (FMOC) protecting group from the linker in the nucleotide complex. The deprotected nucleotide complex was then washed with excess DCM to remove unreacted DBU and form a deprotected nucleotide complex with a free amino group.

b. Addition of Amino Acid

[0209] One equivalent of deprotected nucleotide complex from Example 5A was admixed with 20 equivalents of protected amino acid in dimethylformamide (DMF) and 20 equivalents of 1-hydroxy-benzotriazole (HOBt) under inert atmosphere for 0.5 to 1 hour at room temperature. This reaction condition couples the carboxy-terminus of the amino acid via its pentafluorophenyl ester to the free amino group of the nucleotide complex to form a nucleotide/amino acid

conjugate (the conjugate). The conjugate was then washed in excess DCM to remove unreacted HOBt and the precursor amino acid. The protected amino acid is one of those as described in Example 1, having FMOC and Opfp at the amino and carboxy terminus, and if needed, a blocking group on the side chain as described before.

c. Coupling of Amino Acid to Bifunctional CPG

[0210] As a demonstration of coupling chemistry for addition of an amino acid residue to the bifunctional CPG support, we have synthesized a peptide, H-HPQFVS-aho, and coupled the peptide to bf-CPG. To that end, bf-CPG (428 mg, 26 µmol) was placed in a filter funnel hooked up in a agitating device of an device. Reagents used were the commercially available Fmoc amino acid derivatives either as O-pfp esters (for Val and His[Fmoc]) or the free acid acivated with pyBOP (for Phe, Gln and Pro). Coupling was conducted in the automated peptide synthesizer using conventional chemistry. The success of the coupling reaction was monitored by a positive and a negative Kaiser test after the Fmoc deprotection and the coupling steps, respectively.

[0211] The peptide was deprotected and released from an aliquot of the bf-CPG by treatment with *i.* piperidine in DMF (2/8, 2 x 5 minutes) *ii.* conc. aq. ammonia (shaking for 24 hours). The peptide was analyzed by reversed phase HPLC as one major peak. MS (Ion Spray, *m/e*) 813 (calcd for $C_{39}H_{60}N_{10}O_9$ + H 813).

6. Elongation of the Conjugate

[0212] The conjugate can be lengthened by alternating cycles of addition of nucleotides and amino acids. The following alternating cycles are repeated until the conjugate has desired length amino acid polymer and oligonucleotide polymer.

a. Addition of Nucleotides

[0213] To couple an additional nucleotide, the 5'-OH on the terminal nucleotide is deprotected with DCA following the protocol described previously in Example 4A for the deprotection of the teflon linker/support. Thereafter, a protected nucleotide is added as described in Example 4B.

b. Addition of Amino Acids

[0214] To couple an additional amino acid, the amino-FMOC terminus of the last amino acid added to the conjugate is deprotected with DBU as described previously in Example 5A. Thereafter, a protected amino acid is added as described in Example 5B.

[0215] The cycle of Steps 1 and 2 above adding alternate nucleotides and amino acids can be repeated until the conjugate has polymers of the desired length and structure.

c. Assembly of Polypeptide-Oligonucleotide CPG Conjugates

[0216] As a demonstration of coupling chemistry for preparation of a peptide-oligonucleotide conjugate on a bifunctional CPG support, we have synthesized and coupled an oligodeoxyribonucleotide onto bf-CPG. To that end, oligonucleotides (SEQ ID NO _____: TCT CTC TCT AAA CTC GGG TCT CTC[1]; and SEQ ID NO _____: AGC TAC TTC CCA AGG ATC ACC ACA CTA GCG GGG CCC TAT TCT TAG[2]) were synthesized on a ABI 394 DNA synthesizer using the standard 1 µmole scale synthesis cycle and commercially available reagents and phosphoramidites according to standard synthetic chemistries. The solid support used was the Fmoc-(Fmoc)HPQFVS(DMT)-aho-CPG (20 mg, 1 µmol) produced above in Example _____ that was placed in a commercially available empty synthesis column for the ABI synthesizers. Repetitive yield were calculated to respectively 97.9% (24-mer) and 98.3% (45-mer) as judged from collecting the detritylation, deluting in *p*-toluene sulfonic acid monohydrate in acetonitril (0.1 M) and measuring the absorptions at 498 nm.

[0217] The 45-mer oligonucleotide above has two flanking 15-mer regions that provide a site for PCR primers, and a middle 15-mer that is the coding sequence for the peptide using (arbitrarily) the commaless codons CTA for F, ATC for H, ACC for P, ACA for Q and GCG for V.

[0218] The peptide-oligonucleotide conjugates formed were realesed from the support by conc. aq. ammonia (shaking for more than 24 hours). The peptide-oligonucleotide conjugates were isolated using commercially available

[1] 24-Mer test sequence

[2] 45-Mer sequence. The two flanking 15-mers are PCR primers and the middle 15-mer is the coding sequence for the peptide using (abitrarily) the commaless codons CTA for F, ATC for H, ACC for P, ACA for Q and GCG for V.

OPC-cartridges and analyzed by HPLC and PAGE (radiolabeled with ^{32}P-g-ATP and T4-kinase).

d. Synthesis of a Peptide-Oligonucleotide Conjugate Library on a bf-CPG Support

[0219] Using a bifunctional (bf) CPG support as described in Example 3B, a peptide-oligonucleotide conjugate library is constructed as described below in the following steps:

Step 1. The solid support, bf-CPG, is placed in a commercially available empty synthesis column for use in an ABI 394 DNA synthesizer (Applied Biosystems, Inc., Foster City, CA). The 3'- oligonucleotide primer sequence (SEQ ID NO _____) is synthesized and then chemically attached to the bf-CPG support according to standard manufacturer's procedures using the standard synthesis cycles and commercially available reagents and phosphoramidites.
Step 2. The support after Step 2 is divided into a number of aliquots representing the number of different monomer amino acids to be added at library position 1 of the polypeptides to be synthesized. Each Fmoc-amino acid residue Monomer(1) is coupled according to a protocol consistent with the protection group on the monomer. In a typical protocol bf-CPG is placed in a filter funnel hooked up in a agitating device. The reagents used are the commercially available Fmoc amino acid derivatives either as O-pfp esters or the free acid acivated with pyBOP. The functionalized bf-CPG is the treated with piperidine in DMF (2/8, 2 x 5 minutes) and reacted with the monomer, activated as described above. The success of the coupling reaction is monitored by a positive and a negative Kaiser test after the Fmoc deprotection and coupling steps, respectively. Likewise, the absorption at 302 nm from the deprotected Fmoc/piperidine conjugate acts as an indicator that the coupling reaction proceeeds succesfully. The Monomer(1)-functionalized bf-CPG is recovered by filtration and washed with DMF (2 x), dichloromethane (2 x), diethyl ether (2x) and dried in vacuo to form Monomer(1)-conjugated bf-CPG.
Step 3. The oligonucleotide sequence that is the unit identifier for the appropriate Monomer(1) is attached to each of the Monomer(1)-functionalized bf-CPG's by placing each of the different Monomer(1)-functionalized bf-CPG's in separate commercially available empty synthesis columns for the DNA synthesizers. The encoding oligonucleotide sequence [eg., a comaless trinucleotide unit designated Oligo(1)] that corresponds to the species of Monomer(1) present on that Monomer(1)-functionalized bf-CPG is chemically attached to the Monomer(1)-functionalized bf-CPG by 3 consecutive synthesis cycles on the DNA synthesizer as described previously using the standard synthesis cycles and commercially available reagents and phosphoramidites, to form a Monomer(1), Oligo(1)-bf-CPG conjugate.
Step 4. The different Monomer(1)-functionalized bf-CPG's are pooled, mixed and divided into a number of aliquots representing the number of different amino acid residue monomers to be added in the library at position 2. Monomer(2) is attached to the CPG as described under Step 2 for Monomer(1).
Step 5. The oligonucleotide sequence that is the unit identifier for the appropriate Monomer(2) is attached to each of the Monomer(1)-functionalized bf-CPG's by placing each of the different Monomer(1)-functionalized bf-CPG's in separate commercially available empty synthesis columns for the DNA synthesizers.

These steps, Steps 4-5, are repeated until the appropriate number of monomers and corresponding unit identifier oligonucleotides have been incorporated.
Step 6. The combined solid supports are placed in a commercially available empty synthesis column for the DNA synthesizers. The 5'-oligonucleotide primer sequence is chemically attached to the Monomer(n)-Monomer(1)-functionalized bf-CPG as described previously using the standard synthesis cycles and commercially available reagents and phosphoramidites. The final DMT-group is left attached to the oligonucleotide-peptide conjugates to be used as an affinity tag in the purification step.
Step 7. The combinatorial peptide-oligonucleotide library is deprotected and released from the support by treatment first with TBAF for the appropriate time, and then with conc. aq. ammonia (shaking for more than 24 hours). The peptide-oligonucleotide conjugates were isolated and purified using commercially available OPC-cartridges

7. Removal of Protecting Groups

[0220] After complete synthesis of one or more bifunctional molecules, the protecting groups are removed from the terminal nucleotide, from the terminal amino acid, and from the side chains of protected amino acids.

a. Removal of Nucleotide Protecting Group

[0221] The DMT protecting group on the 5'-OH of the last nucleotide of the oligonucleotide polymer is removed with DCA following the protocol described previously in Example 4A.

b. Removal of Amino Acid Protecting Group

[0222] The FMOC protecting group on the amino-terminus of the last amino acid of the amino acid polymer is removed with DBU as described previously in Example 5A.

c. Removal of Amino Acid Side Chain Protecting

[0223] Conditions for removal of an amino acid side chain protecting group depends on the particular protecting group as follows:

i. Removal of TBS and TSE ester Groups

[0224] One unit of the conjugate is admixed with about 20 equivalents of tetrabutylammonium fluoride (TBAF) in DCM and maintained at room temperature under inert atmosphere overnight to remove the TBS or TMSE ethers protecting the side chains of tyrosine, aspartic acid, glutamic acid, serine, and threonine.

ii. Removal of the Bz Group

[0225] The conjugate is admixed with an excess of aqueous ammonia and maintained at 60°C overnight under inert atmosphere to remove the benzyl (Bz) group protecting the side chain amino group of lysine.

iii. Removal of the MTr, Bum and Tpm Groups

[0226] The conjugate is admixed with 20 to 50 percent TFA and maintained at room temperature for about 5 minutes under inert atmosphere to remove the MTr, Bum or Tpm groups protecting the side chains of arginine, histidine or cysteine, respectively. Thereafter, the conjugate is neutralized with triethanolamine and $CHCl_3$.

iv. Removal of the Formyl Group

[0227] The conjugate is admixed with aqueous buffer at pH 12 and maintained at room temperature for about 5 minutes under inert atmosphere to remove the formyl group protecting the reactive 2-amino group of tryptophan.

8. Cleavage of Conjugate from Solid Support

a. Cleavage of the Teflon Support

[0228] After the protecting groups are removed from the conjugate on the teflon support, the bifunctional molecule is removed from the solid support by admixing the conjugate with a cleaving solution of 100 mM sodium periodate, 100 mM sodium phosphate buffer, pH 7.2, in acetonitrile/water (1:4 v/v). The admixture is maintained with agitation at room temperature with exclusion of light. After 4 hours of agitation, the liquid phase removed and the solid support is washed with excess water and methanol. The wash solutions are then removed and 1 umole of solid support are admixed with 50 ul n-propylamine, 100 ul acetonitrile and 400 ul water and maintained at 55°C for 3 hours. Thereafter, the liquid phase is recovered, evaporated to dryness in vacuo, and the dried product is dissolved in acetonitrile/water. The dissolved product is purified using reverse phase HPLC on an EM LiChrospher 100RP-18m 50um column (4x25) HPLC column. The mobile phase A is 95 % 0.1 TEAA buffer (pH 7.0) and 5 % acetonitrile, and mobile phase B is 5 % TEAA buffer (pH 7.0) and 95 % acetonitrile. The gradient is 100 % A for 5 min, 100 % A to 50 % for 50 min, with a flow rate of 1 mL per min. The homogeneous fraction is collected to yield a solution of pure bifunctional molecule.

[0229] The solution is dialyzed as needed to change the buffer of the purified material.

b. Cleavage of the bf-CPG Support

[0230] The peptide-oligonucleotide conjugate coupled to the bf-CPG support can be cleaved in two different locations.

[0231] As shown in Figure 3, an aqueous ammonia reaction cleaves the sarcosine-succinyl linker moiety, thereby releasing the peptide-oligonucleotide conjugate from the solid support, forming a solution phase conjugate.

[0232] As also shown in Figure 3, irradiation with light at about 350 nm will cleave the photoreactive site present on the serine branch-monomer moiety, thereby releasing the polypeptide from the solid support.

[0233] In both cases, the solid support can readily be separated from the released, soluble, material by filtration to

form isolated soluble conjugate or polypeptide, depending upon the cleavage reaction.

9. Preparation of a Library of Bifunctional Molecules

[0234] Using the synthetic procedures of Examples 1-8, the methods for producing a bifunctional molecule are detailed. To form a library of molecules, additional manipulations are required. First, the synthesis is conducted including the steps of aliquoting, adding different units to each aliquot, and pooling the aliquots to sequentially build the library. Second, if desired, the PCR primer binding sites and the unit identifier oligonucleotides can be added as presynthesized blocks rather than added nucleotide by nucleotide.

a. Synthesis of Protected Oligonucleotides

[0235] Using this procedure, PCR primer binding site oligonucleotides P1 and P2 were synthesized having the nucleotide sequences shown in Table 2, but having an DMT at the oligonucleotide's 5' terminus, and having a CNE ester and an amino diisopropyl phosphoramidate at the oligonucleotide's 3' terminus. Similarly, unit identifier oligonucleotides were synthesized for glycine (Gly) and methionine (Met) having 6 nucleotides per unit and having the blocked termini described above. The unit identifier oligonucleotide sequences are shown in Table 2.

TABLE 2

Designation	Oligonucleotide Sequence
P1	5'-GGGCCCTATTCTTAG-3'
P2	5'-AGCTACTTCCCAAGG-3'
Z^{gly}	5'-CTCATG-3'
Z^{met}	5'-ACGGTA-3'

b. Synthesis of a Library

[0236] The synthesis of a prototype library is described where the chemical unit is an amino acid, the alphabet size is 2, being comprised of glycine and methionine, the unit identifier nucleotide sequence is 6 nucleotides in length, and the chemical polymer length is three amino acids in length. A schematic of the process is shown in Figure 2.

[0237] The solid support prepared in Example 2 is coupled to the linker as described in Example 3. For convenience, the solid support-coupled linker molecule is referred to as LINK. Thereafter, protected oligonucleotide P1 is coupled to LINK as described for a single protected nucleotide in Example 4 to form the structure P1-LINK.

[0238] In Step 1, P1-LINK is divided into two aliquots. The first aliquot is subjected to the sequential coupling of the amino acid residue glycine as described in Example 5, and then coupling of the protected oligonucleotide Z^{gly} as described in Example 4 to form the structure CTCATG-P1-LINK-gly. The second aliquot is similarly coupled to add the amino acid methionine and the oligonucleotide Z^{met} to form the structure ACGGTA-P1-LINK-met. The two aliquots are then pooled to form a mixture of the two bifunctional molecules.

[0239] In Step 2, the pool from Step 1 is divided into two aliquots. The first aliquot is subjected to a sequential coupling as before, adding glycine and the oligonucleotide Z^{gly} to form the structures:

CTCATGCTCATG-P1-LINK-gly.gly, and
CTCATGACGGTA-P1-LINK-met.gly.

The second aliquot is subjected to a sequential coupling as before, adding methionine and the oligonucleotide Z^{met} to form the structures:

ACGGTACTCATG-P1-LINK-gly.met, and
ACGGTAACGGTA-P1-LINK-met.met.

The two aliquots are then pooled to form a mixture of the four bifunctional molecules.

[0240] In Step 3, the pool from Step 2 is divided into two aliquots. The first aliquot is subjected to a sequential coupling as before, adding glycine and the oligonucleotide Z^{gly}. Thereafter, protected oligonucleotide P2 is coupled to the growing bifunctional molecules in the pool as described for a single protected nucleotide in Example 4 to form the struc-

tures:

P2CTCATGCTCATGCTCATGP1-LINK-gly.gly.gly,
P2CTCATGCTCATGACGGTAP1-LINK-met.gly.gly,
P2CTCATGACGGTACTCATGP1-LINK-gly.met.gly, and
P2CTCATGACGGTAACGGTAP1-LINK-met.met.gly.

The second aliquot is subjected to a sequential coupling as before, adding methionine and the oligonucleotide Z^{met}. Thereafter, protected oligonucleotide P2 is coupled to the growing bifunctional molecules in the pool as described for a single protected nucleotide in Example 4 to form the structures:

P2ACGGTACTCATGCTCATGP1-LINK-gly.gly.met,
P2ACGGTACTCATGACGGTAP1-LINK-met.gly.met,
P2ACGGTAACGGTACTCATGP1-LINK-gly.met.met, and
P2ACGGTAACGGTAACGGTAP1-LINK-met.met.met.

The two aliquots are then pooled to form a mixture of the eight bifunctional molecules, which is then further divided and subjected to further sequential coupling steps to produce a library according to the invention.

[0241] By increasing the alphabet size one increases the number of aliquots per step.

[0242] The foregoing specification, including the specific embodiments and examples, is intended to be illustrative of the present invention and is not to be taken as limiting.

Claims

1. A bifunctional molecule according to the following formula:

$$(X_n)\text{-}(\ldots)\text{-}X_4\text{-}X_3\text{-}X_2\text{-}X_1\text{-}B\text{-}Z_1\text{-}Z_2\text{-}Z_3\text{-}Z_4\text{-}(\ldots)\text{-}(Z_n)$$

wherein:

X_n, X_1, X_2, X_3, and X_4 are each single amino acid residues that together comprise a polypeptide,

Z_n, Z_1, Z_2, Z_3, and Z_4 are each nucleotide sequences of the same length, the length being from 2 to 10 nucleotides, which correspond to and identify X_n, X_1, X_2, X_3, and X_4 respectively;

B is a linker molecule operatively linked to X_1 and Z_1, and

n is a position identifier for both X and Z having a value from 4 to 50.

2. A bifunctional molecule according to the following formula:

$$(X_n)\text{-}(\ldots)\text{-}X_4\text{-}X_3\text{-}X_2\text{-}X_1\text{-}B\text{-}P1\text{-}Z_1\text{-}Z_2\text{-}Z_3\text{-}Z_4\text{-}(\ldots)\text{-}(Z_n)\text{-}P2$$

wherein:

X_n, X_1, X_2, X_3, X_4, Z_n, Z_1, Z_2, Z_3, Z_4, B and n are as defined in claim 1, and
P1 and P2 are nucleotide sequences that provide PCR primer binding sites adapted to amplify the polymer identifier oligonucleotide.

3. The bifunctional molecule of claim 2 wherein said P1 and P2 each contain a sequence that defines a restriction endonuclease site when present in a PCR-amplified duplex DNA fragment.

4. The bifunctional molecule of claim 3 wherein said restriction sites are located proximal to said polymer identifier oligonucleotide.

5. The bifunctional molecule of claim 4 wherein said restriction endonuclease sites form non-overlapping cohesive termini upon restriction endonuclease cleavage.

6. A library comprising a plurality of species of bifunctional molecules according to claim 1 or 2.

7. The library of claim 6 wherein each of said species of bifunctional molecules in said plurality is present in molar equivalents of from 0.2 to 10.0.

8. A library comprising a plurality of species of bifunctional molecules according to claim 2, wherein P1 and P2 each have a nucleotide sequence shared by all bifunctional molecule species in the library.

9. A method for identifying a polypeptide polymer that participates in a preselected binding interaction with a biologically active molecule to form a binding reaction complex, the method comprising the steps of:

a) providing a library of bifunctional molecules as described in claim 6; then

b) admixing in solution said library of bifunctional molecules of said step (a) with the biologically active molecule under binding conditions for a time period sufficient to form the binding reaction complex; then

c) isolating the binding reaction complex formed in said step (b); then

d) amplifying the identifier oligonucleotide of the bifunctional molecule within the binding reaction complex isolated in said step (c) by means of PCR; and then

e) sequencing and decoding the identifier oligonucleotide amplified in said step (d) for identifying the polypeptide that participated in the preselected binding interaction.

10. The method of claim 9 wherein said biologically active molecule is affixed to a solid support.

11. The method of claim 9 wherein said biologically active molecule is operatively linked to a binding means capable of binding a binder molecule.

12. The method of claim 11 wherein said binding means is selected from the group consisting of biotin, protein A and magnetic beads.

13. The method of claim 9 wherein said step (d) comprises the following substeps:

i) amplifying the identifier oligonucleotide isolated in said step (C) by a polymerase chain reaction (PCR); and then

ii) sequencing and decoding the PCR amplification product of said substep (I) for identifying the polypeptide that participated in the preselected binding interaction.

14. A method for synthesizing the library of bifunctional molecules in claim 6 wherein the polypeptide of each bifunctional molecule has a length of at least p amino acid residues, where $50 \geq p \geq 4$, and wherein the amino acid residues which comprise the polypeptides are drawn from m species, where $50 \geq m \geq 4$, the method comprising the following steps:

a) dispensing equal aliquots of a nascent linker molecule represented by the following formula:

$$A'\text{-}B\text{-}C'$$

into each of m reaction vessels wherein:

A' is adapted for reaction with any of m activated amino acid residues represented by $(X_1)_m'$; and

C' is adapted for reaction with any of m activated unit identifier nucleotide sequences represented by $(Z_1)_m'$; then

b) dispensing into each of the m reaction vessels of said Step (a) a corresponding aliquot of the m^{th} activated amino acid residue represented by $(X_1)_m'$ and a corresponding aliquot of the m^{th} activated unit identifier nucle-

otide sequence represented by $(Z_1)_m'$ for producing a product nascent bifunctional molecule represented by $(X_1)_m$-B-$(Z_1)_m$; then

c) combining all of the nacsent bifunctional molecules from all m reaction vessels produced in said step (b) for producing an admixture of nascent bifunctional molecules; then

d) dispensing equal aliquots of the admixture of nascent bifunctional molecules from the immediately prior step into each of m reaction vessels; then

e) dispensing into each of the m reaction vessels of said Step (d) a corresponding aliquot of the m^{th} activated amino acid residue represented by $(X_n)_m'$ and a corresponding aliquot of the m^{th} activated unit identifier nucleotide sequence represented by $(Z_n)_m'$ for producing a elongated nascent bifunctional molecule represented by $(X_n)_m$-(...)-$(X_1)_m$-B-$(Z_1)_m$(...)-$(Z_n)_m$, wherein $p \geq n \geq 2$; then

f) combining all of the elongated nacsent bifunctional molecules from all m reaction vessels of step (e) for producing an admixture of elongated nascent bifunctional molecules; and then

g) repeating said steps d-f until $n=p$ and the synthesis of the library is complete.

15. The method of claim 14 wherein said linker molecule is a bifunctional solid support selected from the group consisting of bf-CPG or *o*-NB-bf-CPG.

16. The method of claim 14 wherein said linker molecule is a bifunctional solid support selected from the group consisting of bf-CPG or *o*-NB-bf-CPG.

17. An element of a library of oligopeptide/oligonucleotide conjugates comprising:

a solid support, said solid support being of a type which is dispersible in aqueous solution,

a first linkage unit coupled to said solid support,

a second linkage unit coupled to said first linkage unit,

a bifunctional unit coupled to said second linkage unit,

an oligopeptide attached to said bifunctional unit, and

an oligonucleotide attached to said bifunctional unit, said oligonucleotide encoding said oligopeptide.

18. An element of a library of oligopeptide/oligonucleotide conjugates as described in claim 17 further comprising a cleavable bond for coupling said first linkage unit to said second linkage unit, said cleavable bond being cleavable by exposure to concentrated aqueous ammonia.

19. An element of a library of oligopeptide/oligonucleotide conjugates as described in claim 18 further comprising a cleavable bond for coupling said bifunctional unit to said oligopeptide, said cleavable bond being cleavable by exposure to ultraviolet light.

20. An element of a library of oligopeptide/oligonucleotide conjugates comprising:

a bifunctional unit,

an oligopeptide attached to said bifunctional unit, and

an oligonucleotide attached to said bifunctional unit, said oligonucleotide comprising triplet codons which encode said oligopeptide.

21. An element of a library of oligopeptide/oligonucleotide conjugates as described in claim 20 further comprising a cleavable bond for coupling said bifunctional unit to said oligopeptide, said cleavable bond being cleavable by expo-

sure to ultraviolet light.

Patentansprüche

1. Ein bifunktionales Molekül entsprechend der folgenden Formel:

$$(X_n)—(...)—X_4—X_3—X_2—X_1—B—Z_1—Z_2—Z_3—Z_4—(...)—(Z_n)$$

worin:

X_n, X_1, X_2, X_3 und X_4 jeweils einzelne Aminosäurereste sind, die zusammen ein Polypeptid umfassen, Z_n, Z_1, Z_2, Z_3 und Z_4 jeweils Nukleotidsequenzen der gleichen Länge sind, wobei die Länge 2—10 Nukleotide beträgt, welche zu X_n, X_1, X_2, X_3 bzw. X_4 korrespondieren und sie identifizieren,
B ein Linker-Molekül ist, das operativ an X_1 und Z_1 gebunden ist und
n ein Positionsidentifizierungsmittel für sowohl X wie Z ist mit einem Wert von 4 bis 50.

2. Ein bifunktionales Molekül entsprechend der folgenden Formel:

$$(X_n)—(...)—X_4—X_3—X_2—X_1—B—P1—Z_1—Z_2—Z_3—Z_4—(...)—(Z_n)—P2$$

worin:

X_n, X_1, X_2, X_3, X_4, Z_n, Z_1, Z_2, Z_3, Z_4, B und n wie in Anspruch 1 definiert sind und P1 und P2 Nukleotidsequenzen sind, die PCR-Primerbindungsstellen bereitstellen, die angepasst sind, um das Polymer-Identifikations-Oligonukleotid zu amplifizieren.

3. Das bifunktionale Molekül nach Anspruch 2, worin P1 und P2 jeweils eine Sequenz enthalten, die einen Restriktions-Endonuklease-Stelle definiert, wenn diese in einem PCR-amplifizierten Duplex-DNA-Fragment vorhanden ist.

4. Das bifunktionale Molekül gemäss Anspruch 3, worin die Restriktions-Stellen proximal zum Polymer-Identifikations-Oligonukleotid lokalisiert sind.

5. Das bifunktionale Molekül aus Anspruch 4, worin die Restriktions-Endonuklease-Stellen nichtüberlappende kohäsive Enden nach der Restriktions-Endonuklease-Spaltung bilden.

6. Eine Bibliothek, die eine Vielzahl von Arten bifunktionaler Moleküle nach Anspruch 1 oder 2 umfasst.

7. Die Bibliothek gemäss Anspruch 6, worin jede der Arten bifunktionaler Moleküle in der Vielzahl in molaren Äquivalenten von 0,2 bis 10,0 vorhanden ist.

8. Eine Bibliothek, die eine Vielzahl von Arten bifunktionaler Moleküle nach Anspruch 2 umfasst, worin P1 und P2 jeweils eine Nukleotidsequenz besitzen, die von allen bifunktionalen Molekülarten in der Bibliothek geteilt wird.

9. Ein Verfahren zur Identifikation eines Polypeptid-Polymers, das an einer vorausgewählten Bindungswechselwirkung mit einem biologisch aktiven Molekül teilnimmt, um einen Bindungsreaktionskomplex zu bilden, wobei das Verfahren die folgenden Schritte umfasst;

   a) Bereitstellung einer Bibliothek von bifunktionalen Molekülen, wie in Anspruch 6 beschrieben; dann

   b) Mischung der Bibliothek von bifunktionalen Molekülen aus dem Schritt (a) in Lösung mit dem biologisch aktiven Molekül unter Bindungsbedingungen für einen Zeitraum, der zur Bildung des Bindungsreaktionskomplexes ausreicht; dann

   c) Isolierung des in Schritt (b) gebildeten Bindungsreaktionskomplexes; dann

   d) Amplifikation des Identifikations-Oligonukleotids des bifunktionalen Moleküls in dem Bindungsreaktionskomplex, der in Schritt (c) mit Hilfe von PCR isoliert wurde; und dann

e) Sequenzierung und Dekodierung des Identifikations-Oligonukleotids, das in Schritt (d) amplifiziert wurde, um das Polypeptid, das an der vorausgewählten Bindungswechselwirkung teilnahm, zu identifizieren.

10. Das Verfahren gemäss Anspruch 9, worin das biologisch aktive Molekül an einen festen Träger fixiert ist.

11. Das Verfahren gemäss Anspruch 9, worin das biologisch aktive Molekül operativ an ein Bindungsmittel gebunden wird, welches in der Lage ist, ein Bindungsmolekül zu binden.

12. Das Verfahren gemäss Anspruch 11, worin das Bindungsmittel aus einer Gruppe, welche aus Biotin, Protein A und magnetische Kügelchen besteht, ausgewählt wird.

13. Das Verfahren gemäss Anspruch 9, worin der Schritt (d) die folgenden Unterschritte umfasst:

i) Amplifikation des Identifikations-Oligonukleotids, das in Schritt (c) durch eine Polymerase-Kettenreaktion (PCR) isoliert wird; und dann

ii) Sequenzierung und Dekodierung des PCR-Amplifikationsproduktes von dem Unterschritt (i), um das Polypeptid, das an der vorausgewählten Bindungswechselwirkung teilgenommen hat, zu identifizieren.

14. Ein Verfahren, um die Bibliothek von bifunktionalen Molekülen aus Anspruch 6 zu synthetisieren, worin das Polypeptid von jedem bifunktionalen Molekül eine Länge von mindestens p Aminosäureresten aufweist, wobei $50 \geq p \geq 4$ ist und worin die Aminosäurereste, die die Polypeptide umfassen, aus m Arten entnommen werden, wobei $50 \geq m \geq 4$ ist, wobei das Verfahren die folgenden Schritte umfasst;

a) Ausgabe gleicher Aliquots eines Start-Linker-Moleküls, das durch die folgende Formel dargestellt wird:

$$A'—B—C'$$

in jeweils eines von m Reaktionsgefässen, wobei:

A' für die Reaktion mit jedem von m aktivierten Aminosäureresten, die durch $(X_1)_m'$ dargestellt werden, angepasst ist; und
C' für die Reaktion mit jedem von m aktivierten Einheits-Identifikations-Nukleotidsequenzen, die durch die Formel $(Z_1)_m'$ dargestellt werden, angepasst ist; dann

b) Ausgabe eines korrespondierenden Aliquots des m-ten aktivierten Aminosäurerestes, der durch $(X_1)_m'$ dargestellt wird und eines korrespondierenden Aliquots der m-ten aktivierten Einheits-Identifikations-Nukleotidsequenz, die durch $(Z_1)_m'$ dargestellt wird, in jedes der m Reaktionsgefässe aus Schritt (a), um ein bifunktionales Starter-Molekül-Produkt, das durch $(X_1)_m—B—(Z_1)_m$ dargestellt wird, herzustellen; dann

c) Kombination aller bifunktionaler Starter-Moleküle aus allen m Reaktionsgefäßen, die in Schritt (b) hergestellt wurden, um eine Mischung von bifunktionalen Starter-Molekülen herzustellen; dann

d) Ausgabe gleicher Aliquots der Mischung von bifunktionalen Molekülen aus dem gerade vorausgegangenen Schritt in jedes der m Reaktionsgefäße; dann

e) Ausgabe eines korrespondierenden Aliquots des m-ten aktivieren Aminosäurerestes, der durch $(X_n)_m'$ dargestellt wird, und eines korrespondierenden Aliquots der m-ten aktivierten Einheits-Identifikations-Nukleotidsequenz, die durch $(Z_n)_m'$ dargestellt wird, in jedes der m Reaktionsgefässe aus Schritt (d), um ein verlängertes bifunktionales Starter-Molekül herzustellen, das durch $(X_n)_m—(...)—(X_1)_m—B—(Z_1)_m—(...)—(Z_n)_m$ dargestellt wird, worin $p \geq n \geq 2$ ist; dann

f) Kombination aller verlängerten bifunktionalen Starter-Moleküle von allen m Reaktionsgefässen aus Schritt (e), um eine Mischung verlängerter bifunktionaler Starter-Moleküle herzustellen; und dann

g) Wiederholung der Schritte d—f bis n=p und die Synthese der Bibliothek vollständig ist.

15. Das Verfahren gemäss Anspruch 14, worin das Linker-Molekül ein bifunktionaler fester Träger ist, der aus der

Gruppe, die aus bf—CPG oder o—NB—bf—CPG besteht, ausgewählt wird.

16. Das Verfahren gemäss Anspruch 14, worin das Linker-Molekül ein bifunktionaler fester Träger ist, der aus der Gruppe, die aus bf—CPG oder o—NB—bf—GPG besteht, ausgewählt wird.

17. Ein Element aus einer Bibliothek von Oligopeptid-/Oligonukleotid-Konjugaten, umfassend:

    einen festen Träger, wobei der Träger von einer solchen Art ist, welche in wässriger Lösung dispergierbar ist,
    eine erste Bindungseinheit, die an dem festen Träger gekoppelt ist,
    eine zweite Bindungseinheit, die an die erste Bindungseinheit gekoppelt ist,
    eine bifunktionale Einheit, die an die zweite Bindungseinheit gekoppelt ist,
    ein Oligopeptid, das an der bifunktionalen Einheit angeheftet ist, und
    ein Oligonukleotid, das an der bifunktionalen Einheit angeheftet ist, wobei das Oligonukleotid das Oligopeptid kodiert.

18. Ein Element einer Bibliothek von Oligopeptid—/Oligonukleotid—Konjugaten, wie in Anspruch 17 beschrieben, welches desweiteren eine spaltungsfähige Bindung für die Kopplung der ersten Bindungseinheit an die zweite Bindungseinheit umfasst, wobei die spaltungsfähige Bindung gespalten werden kann, wenn sie konzentriertem wässrigen Ammoniak ausgesetzt wird.

19. Ein Element einer Bibliothek von Oligopeptid—/Oligonukleotid—Konjugaten, wie in Anspruch 18 beschrieben, welches desweiteren eine spaltungsfähige Bindung für die Kopplung der bifunktionalen Einheit an das Oligopeptid umfasst, wobei die spaltungsfähige Bindung gespalten werden kann, wenn sie ultraviolettem Licht ausgesetzt wird.

20. Ein Element einer Bibliothek von Oligopeptid—/Oligonukleotid—Konjugaten, umfassend:

    eine bifunktionale Einheit
    ein Oligopeptid, das an der bifunktionalen Einheit angeheftet ist, und
    ein Oligonukleotid, das an der bifunktionalen Einheit angeheftet ist, wobei das Oligonukleotid Triplett-Codons umfasst, welche das Oligopeptid kodieren.

21. Ein Element einer Bibliothek von Oligopeptid—/Oligonukleotid—Konjugaten, wie in Anspruch 20 beschrieben, welches desweiteren eine spaltungsfähige Bindung für die Kopplung der bifunktionalen Einheit an das Oligopeptid, wobei die spaltungsfähige Bindung gespalten werden kann, wenn sie ultraviolettem Licht ausgesetzt wird.

Revendications

1. Molécule bifonctionnelle selon la formule suivante :

$$(X_n)\text{-}(\ldots)\text{-}X_4\text{-}X_3\text{-}X_2\text{-}X_1\text{-}B\text{-}Z_1\text{-}Z_2\text{-}Z_3\text{-}Z_4\text{-}(\ldots)\text{-}(Z_n)$$

   où

    X_n, X_1, X_2, X_3, et X_4 sont chacun des résidus d'acides aminés qui ensemble forment un polypeptide,
    Z_n, Z_1, Z_2, Z_3, et Z_4 sont chacun des séquences de nucléotides de la même longueur, la longueur étant de 2 à 10 nucléotides, qui correspondent à une identité X_n, X_1, X_2, X_3, et X_4 respectivement;
    B est une molécule de linker opérativement enchaînée à X_1 et Z_1, et
    n est un identificateur de position pour X et Z ayant une valeur de 4 à 50.

2. Molécule bifonctionnelle selon la formule suivante

$$(X_n)\text{-}(\ldots)\text{-}X_4\text{-}X_3\text{-}X_2\text{-}X_1\text{-}B\text{-}P1\text{-}Z_1\text{-}Z_2\text{-}Z_3\text{-}Z_4\text{-}(\ldots)\text{-}(Z_n)\text{-}P2$$

   où

    X_n, X_1, X_2, X_3, X_4, Z_n, Z_1, Z_2, Z_3, Z_4, B et n sont tels que définis à la revendication 1 et
    P1 et P2 sont des séquences de nucléotides qui produisent des sites de liaison d'amorce de PCR adaptés à amplifier l'oligonucléotide identificateur du polymère.

3. Molécule bifonctionnelle de la revendication 2, où chacun de P1 et P2 contient une séquence qui définit un site d'endonucléase de restriction lors d'une présence dans un fragment d'ADN en duplex amplifié par PCR.

4. Molécule bifonctionnelle de la revendication 3, où lesdits sites de restriction sont placés près dudit oligonucléotide identificateur du polymère.

5. Molécule bifonctionnelle de la revendication 4, où lesdits sites d'endonucléase de restriction forment des extrémités cohésives ne se recouvrant lors d'une scission par l'endonucléase de restriction.

6. Banque comprenant un certain nombre d'espèces de molécules bifonctionnelles selon la revendication 1 ou 2.

7. Banque de la revendication 6, où chacune desdites espèces de molécules bifonctionnelles dans ladite quantité est présente à des équivalents molaires de 0,2 à 10,0.

8. Banque comprenant un certain nombre d'espèces de molécules bifonctionnelles selon la revendication 2, où chacun de P1 et P2 a une séquence de nucléotides partagée par toutes les espèces de molécules bifonctionnelles dans la banque.

9. Méthode pour identifier un polymère polypeptidique qui participe dans une interaction de liaison présélectionnée avec une molécule biologiquement active pour former un complexe de réaction de liaison, la méthode comprenant les étapes de :

   a) produire une banque de molécules bifonctionnelles telle que décrite à la revendication 6 ; puis
   b) mélanger en solution ladite banque de molécules bifonctionnelles de ladite étape (a) avec la molécule biologiquement active en conditions de liaison pendant une période de temps suffisante pour former le complexe de réaction de liaison ; puis
   c) isoler le complexe de réaction de liaison formé dans l'étape (b) ; puis
   d) amplifier l'oligonucléotide identificateur de la molécule bifonctionnelle dans le complexe de réaction de liaison isolé dans ladite étape (c) par PCR ; et puis
   e) mettre en séquence et décoder l'oligonucléotide identificateur amplifié dans ladite étape (d) pour identifier le polypeptide qui a participé à l'interaction de liaison présélectionnée.

10. Méthode de la revendication 9, où ladite molécule biologiquement active est fixée à un support solide.

11. Méthode de la revendication 9, où ladite molécule biologiquement active est activement enchaînée à un moyen de liaison capable de lier une molécule de liant.

12. Méthode de la revendication 11, où ledit moyen de liaison est sélectionné dans le groupe consistant en biotine, protéine A et perles magnétiques.

13. Méthode de la revendication 9, où ladite étape (d) comprend les sous-étapes suivantes :

   i) amplifier l'oligonucléotide identificateur isolé à ladite étape (c) par une réaction en chaîne de polymérase (PCR) ; et puis
   ii) mettre en séquence et décoder le produit d'amplification par PCR de ladite sous-étape (I) pour identifier le polypeptide qui a participé à l'interaction de liaison présélectionnée.

14. Méthode pour la synthèse de la banque de molécules bifonctionnelles de la revendication 6, où le polypeptide de chaque molécule bifonctionnelle a une longueur d'au moins p résidus d'acides aminés, où $50 \geq p \geq 4$, et où les résidus d'acides aminés qui forment les polypeptides sont prélevés de m espèces, où $50 \geq m \geq 4$, la méthode comprenant les étapes suivantes :

   a) distribuer des aliquotes égales d'une molécule linker naissante représentée par la formule suivante :

$$A'\text{-}B\text{-}C'$$

   dans chacun des m récipients réactionnels où :

A' est adapté à une réaction avec chacun des m résidus d'acides aminés activés représentés par $(X_1)_m$' ; et

C' est adapté à une réaction avec chacune des m séquences de nucléotides d'identificateur unitaire activées représentées par $(Z_1)_m$' ; puis

b) distribuer dans chacun des m récipients réactionnels de ladite étape (a), une aliquote correspondante du mième résidu d'acide aminé activé représenté par $(X_1)_m$' et une aliquote correspondante de la mième séquence de nucléotides d'identificateur activé représentée par $(Z_1)_m$' pour donner une molécule bifonctionnelle naissante produite représentée par $(X_1)_m$-B-$(Z_1)_m$' ; puis
c) combiner toutes les molécules bifonctionnelles naissantes de tous les m récipients réactionnels produits à l'étape (b) pour produire un mélange de molécules bifonctionnelles naissantes ; puis
d) distribuer des aliquotes égales du mélange des molécules bifonctionnelles naissantes de l'étape immédiatement antérieure dans chacun des m récipients réactionnels ; puis
e) distribuer, dans chacun des m récipients réactionnels de ladite étape (d), une aliquote correspondante du mième résidu d'acide aminé activé représenté par $(X_n)_m$' et une aliquote correspondante de la mième séquence de nucléotides d'identificateur activé représentée par $(Z_n)_m$' pour produire une molécule bifonctionnelle naissante allongée représentée par $(X_n)_m$-(...)-$(X_1)_m$-B-$(Z_1)_m$(...)-$(Z_n)_m$, où $p \geq n \geq 2$; puis
f) combiner toutes les molécules bifonctionnelles naissantes allongées de m récipients réactionnels de l'étape (e) pour produire un mélange de molécules bifonctionnelles naissantes allongées ; et puis
g) répéter lesdites étapes d-f jusqu'à ce que n=p et que la synthèse de la banque soit terminée.

15. Méthode de la revendication 14, où ladite molécule linker est un support solide bifonctionnel sélectionné dans le groupe consistant en bf-CPG ou o-NB-bf-CPG.

16. Méthode de la revendication 14, où ladite molécule linker est un support solide bifonctionnel sélectionné dans le groupe consistant en bf-CPG ou o-NB-bf-CPG.

17. Elément d'une banque de conjugués oligopeptides/oligonucléotides comprenant :

un support solide, ledit support solide étant d'un type qui est dispersible dans une solution aqueuse,
une première unité d'enchaînement couplée audit support solide,
une seconde unité d'enchaînement couplée à ladite première unité d'enchaînement,
une unité bifonctionnelle couplée à ladite seconde unité d'enchaînement,
un oligopeptide attaché à ladite unité bifonctionnelle, et
un oligonucléotide attaché à ladite unité bifonctionnelle, ledit oligonucléotide codant pour ledit oligopeptide.

18. Elément d'une banque de conjugués oligopeptides/oligonucléotides comme décrit à la revendication 17, comprenant de plus une liaison scindable pour couplage de ladite première unité d'enchaînement à ladite seconde unité d'enchaînement, ladite liaison scindable étant scindable par exposition à de l'ammoniaque aqueuse concentrée.

19. Elément d'une banque de conjugués oligopeptides/oligonucléotides comme décrit à la revendication 18, comprenant de plus une liaison scindable pour couplage de ladite unité bifonctionnelle audit oligopeptide, ladite liaison scindable étant scindable par exposition à une lumière ultraviolette.

20. Elément d'une banque de conjugués oligopeptides/oligonucléotides, comprenant

une unité bifonctionnelle,
un oligopeptide attaché à ladite unité bifonctionnelle, et
un oligonucléotide attaché à ladite unité bifonctionnelle, ledit oligonucléotide comprenant des codons en triplet qui codent pour ledit oligopeptide.

21. Elément d'une banque de conjugués oligopeptides/oligonucléotides comme décrit à la revendication 20, comprenant de plus une liaison scindable pour coupler ladite unité bifonctionnelle audit oligopeptide, ladite liaison scindable étant scindable par exposition à de la lumière ultraviolette.

```
                 Sty I                       Apa I
5' AGCTACTTCCCAAGG[coding sequence   ]GGGCCCTATTCTTAG 3'
3' TCGATGAAGGGTTCC[anticoding strand]CCCGGGATAAGAATC 5'

                    |  Step 1   Cleavage by
                    |           Sty I & Apa I
                    v

5' AGCTACTTCC      CAAGG[coding sequence   ]GGGCC      CTATTCTTAG 3'
3' TCGATGAAGGGTTC      C[anticoding strand]C      CCGGGATAAGAATC 5'

                    |  Step 2   Biotinylation
                    v

5' AGCTACTTCCC    CAAGG[coding sequence   ]GGGCC      CTATTCTTAG 3'
3' TCGATGAAGGGTTC  BBCC[anticoding strand]C      CCGGGATAAGAATC 5'
```

FIG. I

P1 - LINK

↓ Step 1

CTCATG-P1-LINK-gly

ACGGTA-P1-LINK-met

↓ Step 2

CTCATGCTCATG-P1-LINK-gly.gly
CTCATGACGGTA-P1-LINK-met.gly

ACGGTACTCATG-P1-LINK-gly.met
ACGGTAACGGTA-P1-LINK-met.met

↓ Step 3

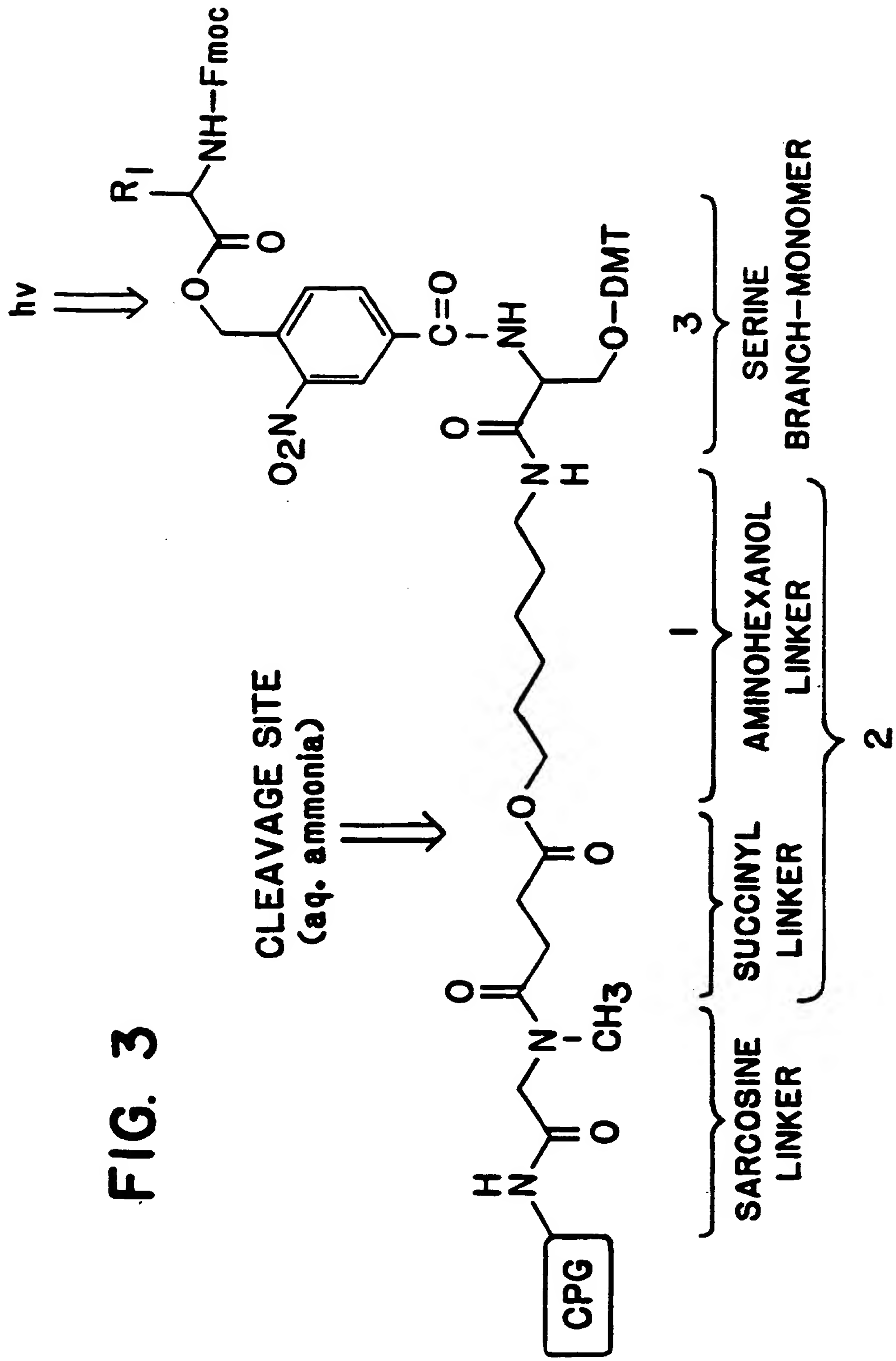
P2CTCATGCTCATGCTCATGP1-LINK-gly.gly.gly
P2CTCATGCTCATGACGGTAP1-LINK-met.gly.gly
P2CTCATGACGGTACTCATGP1-LINK-gly.met.gly
P2CTCATGACGGTAACGGTAP1-LINK-met.met.gly

P2ACGGTACTCATGCTCATGP1-LINK-gly.gly.met
P2ACGGTACTCATGACGGTAP1-LINK-met.gly.met
P2ACGGTAACGGTACTCATGP1-LINK-gly.met.met
P2ACGGTAACGGTAACGGTAP1-LINK-met.met.met

P1 = GGGCCCTATTCTTAG
P2 = AGCTACTTCCCAAGG

FIG. 2

FIG. 3



hv
NH—Fmoc
R1
O2N
C=O
NH
O—DMT
CLEAVAGE SITE
(aq. ammonia)
CPG
CH3
SARCOSINE LINKER
SUCCINYL LINKER
AMINOHEXANOL LINKER
SERINE BRANCH—MONOMER
1
2
3